A DOMAIN-CENTRIC APPROACH TO DESIGNING USER INTERFACES OF VIDEO RETRIEVAL SYSTEMS

Dan Edward Albertson II

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements for the degree

Doctor of Philosophy

in the School of Library and Information Science,

Indiana University

July 2007

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Javed Mostafa, Ph. D.

Kiduk Yang, Ph. D.

Youn-Kyung Lim, Ph. D.

Catharine Wyss, Ph. D.

June 21, 2007

Acknowledgment

I would first like to acknowledge the chairperson of my research committee, Dr. Javed Mostafa. Dr. Mostafa's hard work and direction is so greatly appreciated; I could not have achieved this goal without his steadfast support. I also owe Dr. Mostafa a debt of gratitude for allowing me to serve on various research grants that provided stable financial support throughout my entire doctoral study. I give a sincere "Thank You" to Dr. Mostafa, not only for being an extraordinary mentor but also a dear friend. I would also like to acknowledge the other faculty members who served on my research committee, including Dr. Kiduk Yang, Dr. Youn-Kyung Lim, and Dr. Catharine Wyss. Their constructive feedback and continual guidance throughout my doctoral program and dissertation research were exceptional. Next, the School of Library and Information Science (SLIS) at Indiana University, Bloomington (IUB) is acknowledged for providing a superior educational experience at both the master's and doctoral levels, and also offering generous financial support during the earlier years of my doctoral program. My acknowledgement of SLIS is extended to its world-class faculty and great administrative and IT staff. Next, I would especially like to thank my family. First, I deeply value my wife, Tara, and daughter, Lucy Mae, for being so supportive throughout my educational pursuit. I will also never undervalue the encouragement provided by my parents, Dan and Carol, who have been with me every step of the way. I would also like to acknowledge all of my other friends, colleagues, and family members who provided such great insight, advice, and times of relief throughout this process. Finally, thanks to all of the IT professionals around IUB who spent numerous hours assisting me with system development efforts. I am especially indebted to Stephanie Burks at the Research and

Technical Services and the Research Database Complex at UITS. I would like to thank Stephanie for all of her help with implementing and populating multiple databases. Also, thanks to Arvind Gopu at the High Performance Computing Support at UITS for earlier assistance with information processing and other developmental efforts.

Dan Edward Albertson II

A DOMAIN-CENTRIC APPROACH TO DESIGNING USER INTERFACES OF VIDEO RETRIEVAL SYSTEMS

User- and task-centric efforts in video information retrieval (IR) research are needed because current experiments are showing few significant results. It is our belief that unsatisfactory results in video IR can be partially attributed to the overemphasis on technologically-driven approaches to interface development and system evaluation. This study explored variables that have been consistently overlooked in video retrieval efforts, including those related to domain and search tasks. The underlying goal of this study is to promote alternative means for evaluating video retrieval systems, and to make progress toward developing new design principles and a video seeking model. A series of interactive search runs were conducted using a video retrieval system called ViewFinder. ViewFinder was implemented to search and browse the NASA K – 16 Science Education Programs. The system includes new design features that take into account the unique characteristics of the domain and associated tasks. Users with a background in Science Education, including teachers and academic majors, were recruited to perform a number of search tasks. Results from the search experiments were collected and analyzed using both objective and subjective measures. From these results, researchers gained further knowledge about domain-centric video search tasks, including how textual, visual, and hybrid tasks were all deemed important by science educators. Further analysis of experimental results also revealed associations between search tasks, user interaction, interface features and functions, and system effectiveness. The evaluation of individual

interface features and functions exhibited that keyword searching was significant for retrieving Science Education video. However, these experiments also produced positive results for various visual search features. Unlike keyword searching, which was consistent and effective across many task types, the use and effectiveness of visual search and browse features were shown to be task dependent. Overall, the results from this study highlight the importance of user- and task-centric methods in video retrieval, as they provided researchers with additional understanding of the influences of domain-specific search tasks on user interaction with video systems. In addition, the experimental methodology employed for this study encourages future foundations for developing and evaluating video search interfaces designed for specific domains and search tasks.

Table of Contents

List of Tables

List of Figures

List of Images

Chapter 1

Introduction

A survey of video retrieval research shows that there is little difference between current approaches to developing and evaluating user interfaces, and that newer designs consistently lack any significant improvements. This stagnancy in interface development for video retrieval systems is demonstrated by the fact that a majority of user interfaces function on an image level. Furthermore, individual images serve as the main visual feature for retrieving and presenting video information. The complexity[1] of video information makes it reasonable to argue that video retrieval research needs to investigate user interaction and interface development more independently from other areas of information retrieval (IR).

However, video retrieval as an area of research remains relatively monotonous and the mainstream of retrieval systems continues to provide indistinguishable interface features and functions. Some common video search functions include querying by

[1] Text, images, audio, and moving-images can all be automatically extracted from video.

keywords, visual information[2], and any combination of keywords and visual information[3]. Keyword search functions are frequently implemented using video transcripts[4] and bibliographic records. Visual searching uses automatically extracted attributes, including colors, shapes, and textures, to retrieve video information.

The problems associated with video retrieval's regression are only compounded by the fact that a majority of current experiments have shown insignificant impacts[5] for implementing newer and more complex search functions. Studies that evaluate complex search functions, including those that exclusively rely on content-based retrieval, rarely produce results that are superior to keyword searching alone. Consequently, text-based search functions are consistently shown to be the most important for retrieving video information.

Many research laboratories have drawn similar conclusions. The University of North Carolina at Chapel Hill (UNC) recently evaluated different versions of their Open Video system where a variety of search functions were compared. Results from their interactive search experiments did not show any significant differences, in terms of recall and precision, between a hybrid system and a transcript-only search system (Yang, Wildemuth, & Marchionini, 2004). Their hypothesis that the hybrid system would be superior across all search topics was not supported. UNC's results also showed that

---

[2] The use of visual information for retrieving video is also commonly referred to as content-based retrieval.

[3] Combining keywords and visual information for video searching is referred to as hybrid retrieval.

[4] There are different types of video transcripts including automatic speech recognition (ASR) and closed-caption (CC) outputs.

[5] "Impact" or "performance" in video retrieval research is typically measured by recall and precision.

visual searching alone actually decreased retrieval performance when compared to a transcript-only search (Yang et al., 2004).

The TREC Video Retrieval Evaluation Workshop (TRECVID) is a highly regarded research forum that draws participants from both academia and industry. Many experiments performed for participation in the TRECVID search task have produced results that support UNC's findings. The format of the TRECVID search task consists of issuing common datasets[6] and information need statements,[7] and requiring all participants to submit search results from experimental runs[8]. Results from all experimental search runs are then commonly evaluated and issued to TRECVID participants (TREC Video Retrieval Evaluation Homepage., 2000).

Researchers from Dublin City University (DCU) regularly evaluate their Físchlár Digital Video Library using the TRECVID protocol. DCU has participated in TRECVID since its inception and consistently ranks among the best. In its 2004 study, DCU concluded that keyword searching was still the most important function of its retrieval system (Cooke et al., 2005). In the same study, however, DCU claimed to have made progress in content-based retrieval as results showed a limited number of topics where visual searching outperformed keyword searching (Cooke et al., 2005).

The Informedia Project at Carnegie Mellon University (CMU) is another annual participant in TRECVID. In 2004, CMU researchers tested different versions of their Informedia system across multiple search runs (Hauptmann et al., 2005). Half of their

[6] Over the past three years, TRECVID's dataset has included CNN Headline News, ABC World News Tonight, and CSPAN news programming.

[7] Information need statements are referred to as search topics by TRECVID organizers and participants.

[8] By TRECVID standards, a search run is one full search experiment, or performing all search topics once.

search runs were performed using a "full version"[9] of the Informedia system while the other half tested a visual-only search system. Results from the search runs were compared. The full system significantly outperformed the visual system on each of the interactive search runs, even when comparing novice users on the full system to expert users on the visual system (Hauptmann et al., 2005). CMU's results support that textual information is still most important for retrieving video, and visual-only search functions remain ineffective for a majority search topics.

The examples above demonstrate that video retrieval, as a field, is not reaching its potential in the areas of systems development and evaluation. These studies also show that results from video retrieval experiments lack the significance needed for creating standards, and that future efforts have nominal foundations for investigating retrieval problems. It should be noted that the purpose of this study is not to sweepingly discount any particular search functions or interface features. Nor is it the goal of this study to completely reject systems-centered evaluation methods. This study intends to reveal that important means for system evaluation are missing from current video IR research, and that interface development efforts can be strengthened by employing a broader set of experimental methods. Moreover, the authors suggest that not all problems related to interface features and functions[10] exist on a systems level. This study recommends that the scope of many video IR experiments is too narrow, or system-centered, and other

[9] CMU's full system implemented both visual and textual search functions, i.e. a hybrid retrieval system..

[10] Interface features and functions are separate components of a video retrieval system in a sense that features may refer to graphical components of an interface, and functions include system processes that operate behind, i.e. hidden from, the interface. For example, a keyword search field may be an example of an interface feature while natural language search may be an example of a function.

important problems exist within the evaluation process. Many studies, including those that show significant results, are not all that revealing because important variables are frequently overlooked.

Variables that are not being evaluated in current video IR research is one primary focus of this study. The variables at the center of this investigation include those related to video search tasks. Drawing from previous research in text and image IR, the authors believe there is significant evidence to support that exploring problems from the perspective of search tasks can have positive implications for IR studies. What exactly are video search tasks? In general, video search tasks depict what information users are looking for, why they are looking for it, and how they will use it. Video search tasks not only comprise simple information need statements, i.e. TRECVID topics, but also provide insight into user context and application purposes. Examples of video search tasks as identified throughout various domains include:

- A graphic artist looking for video of the ocean that symbolizes anger and rage
- A historian searching for a clip of Charles de Gaulle walking through the streets of Paris
- A criminal investigator looking for surveillance video of a car jacking
- A traveler searching for shots of the U.S. Doppler radar for a weather report
- A medical instructor needing video of a medical scope that shows signs of prostate cancer to show during class

Examples of video search tasks are virtually endless. The search tasks above demonstrate that they span many domains, i.e. professional communities or academic disciplines, and serve many purposes. Video search tasks can be multidimensional, conceptually specific or vague, and have many other defining characteristics.

Evidence of how search tasks are being overlooked in video IR experiments is illustrated by the DCU and CMU experiments described above. In their TRECVID 2004 study, both DCU and CMU claim to have observed a significant impact for visual searching on certain topics, but neither elaborate on the characteristics of those topics, or describe where or how keyword searching is most important. Other studies that have attempted to measure the impact of video search tasks on retrieval performance, including the UNC study, are also insufficient because the topics and evaluation measures were confined to the TRECVID protocol[11]. This study expands upon the DCU, CMU, and UNC studies, and explores video retrieval problems from a more domain and task-specific perspective, and evaluates interface support accordingly.

While search tasks have been explored throughout text and image IR research, little effort has been given to investigating the characteristics and impact of video search tasks. This deficiency in video IR research is important because task characteristics have been shown to influence search performance and user interaction. Findings from task-centric research can be used to support indexing protocols and systems and interface development (Vakkari, 1999).

A research team from the National University of Singapore (NUS) is one of the few TRECVID participants who explored questions related to video search tasks. One aspect of NUS's TRECVID study included automatic query analysis where all search

[11] TRECVID may not be the best forum to investigate task-centered video retrieval research due to the limited scope of the search topics, a generic dataset, and the narrowness of evaluation.

topics were categorized. Topic categories created by NUS included *person, sports, finance, weather, disaster,* and *general* (Chua et al., 2005). NUS researchers hypothesized that they could produce better search results by considering topic category for the retrieval process. NUS's interactive search experiments were designed to evaluate several retrieval algorithms. Results of the NUS study supported their assumptions and the retrieval approach was shown to be beneficial; readers should refer to the original source for results on all the experimental factors (Chua et al., 2005). The importance of this study comes from researchers having recognized that search topics, one aspect a search task, can affect retrieval and benefit system functionality. Another study resembling the NUS experiments was conducted by researchers from CMU where similar findings were published in the *Proceedings of the 2004 ACM Multimedia Conference* (ACM'MM) (Yan, Yang, & Hauptmann, 2004).

The methods used by the NUS study can be one approach for exploring task-centric video retrieval. However, research should expand beyond TRECVID studies and advance video retrieval in a more user-centered direction. One contribution from this study is that problems focus on users, not systems. User-centered studies in video IR are needed because current research is predominately technology driven, and consideration of the user is frequently excluded from developmental processes. Although there has been significant progress in user-centered research in text and image IR, advances in user-centered video IR have been negligible.

User-centered video IR research, independent of text and image IR research, is necessary because users employ different strategies when interacting with and searching for video. Differences among video, text, and image searching can be attributed to the disparities between the makeup and structure of the information formats. Video is highly complex as image, text, and audio are all components of video information, and can all be automatically extracted from it. Video's complexity causes actions associated with searching, i.e. relevance assessment, query formulation, etc., to be approached differently. Investigating the associations between user interactions and various features of video information, from the perspective of search tasks, is another goal of this study. (The specific goals of this study will be comprehensively surveyed in Chapter 2.)

This study also examines why users from a particular domain search for and use digital video. Exploring these motivations can help in the discovery of domain-specific tasks for video searching. When exploring problems from the perspective of search tasks, researchers must also consider and address implication for the domain, and vice versa. Domain-centric research, a more focused area of user-centered research, can start by: 1) analyzing users in a particular - video searching - context and 2) identifying the different tasks performed throughout all levels[12] of video retrieval. Researchers can employ domain-centric methods to investigate video search tasks that are unique and important for certain professional or academic domains, and determine how to best support users with searching interfaces.

Investigating video IR from the perspective of search tasks is sure to lead to many more beneficial and practical implications. Most importantly, researchers can formulate guidelines or standards for developing video retrieval systems that support specific tasks. Many different facets of video systems development, including interface design, will benefit from domain-centric studies. Researchers can also begin exploring video-related

[12] Different levels of video searching include text-based, visual, and hybrid searching.

tasks beyond searching, such as those associated with video use and manipulation. This study examines system design principles that allow users to perform a wide range of tasks for vide retrieval.

As previously stated, problems centering on search tasks have been investigated throughout traditional IR studies. Moreover, the concept of task-specific retrieval systems is not new to text IR researchers. Researchers from Northwestern University (NU) claimed they could garner vital information for retrieving relevant documents by automatically interpreting users' interactions (Budzik & Hammond, 2000). NU researchers assert that information seeking is not an accidental or random process, but that user context, a parameter related to domains and tasks, is important for retrieving relevant information. They also maintain that current systems remain ineffective because many do not consider search tasks or application purposes when retrieving documents. NU researchers identify several issues that arise when excluding contextual or task-related information from the retrieval process, including interpretations of relevance, ambiguity of search terms, and varying audiences (Budzik & Hammond, 2000). These issues are claimed to be particularly relevant to web search engines because few search engines collect contextual information, and user queries are typically too vague to portray contexts or search tasks.

To explore these assumptions, NU researchers implemented and tested Watson, an information management assistant (IMA). Watson used its *anticipator* function to interpret users' tasks and predict information needs. Watson was integrated into a widely-used software system, i.e. a word processor, where the *content analyzer* deciphered document type and analyzed document content. Based on input from the *anticipator* and *content analyzer,* search results were continuously filtered, processed, and returned to the user without any explicit queries. Watson also provided a search interface where users could manually query the system as needed and retrieve additional task-specific results. Experiments using Watson returned positive results,[13] and the task-related search functions were shown to be beneficial. The Watson project demonstrates that IR researchers have recognized the importance of search tasks and have been proactive in conducting task-related research. Other studies that explore task-oriented retrieval are found throughout relevance feedback, user profiling, and search term disambiguation research (Budzik & Hammond, 2000).

The Watson project and other experiments previously discussed can all be classified as IR research. Although NU researchers present an interesting study for measuring the impact of search tasks, their experiments were confined to one particular set of interface features and functions. Moreover, findings from other areas of information science, including the exploration of additional domain-specific tasks, were mostly overlooked. Vakkari (1999), on the other hand, analyzed search tasks on a higher level, or in more theoretical terms.

In his study, Vakkari (1999) claimed that systems development should begin with the tasks that need to be supported, as opposed to any particular technology or data. In addition, he states that the main purpose of retrieval systems should reflect the search tasks and user context (Vakkari, 1999). Vakkari bases these claims on findings from highly cited studies, such as Checkland & Holwell, 1998; cf. Belkin, Seeger & Wersig, 1983; Ingwersen, 1992; Kunz, Rittel & Schwuchow1977; Wilson, 1981.

---

[13] Results were deemed satisfactory by the users, and Watson was judged to be a beneficial search tool.

Vakkari (1999) surveyed search tasks on more theoretical grounds. He asserted that in order to understand the true essence of IR, researchers must first understand actions related to search tasks. These "information actions" include search strategies, relevance assessment, and information needs and types (Vakkari, 1999).

Vakkari's survey was not limited to IR literature; an analysis of search tasks as identified throughout information seeking (IS) research was also performed. Vakkari (1999) identified several key ideas that are common among many IS studies including how humans fundamentally seek out information in order to fill a void in knowledge for solving certain problems. He claims that in order to fully understand the information seeking process, i.e. information needs and actions capable of fulfilling them, it is essential to analyze and understand search tasks (Vakkari, 1999). Certain factors surrounding search tasks, including task complexity and problem structure, will influence search strategies, search performance, and information needs.

Vakkari's study is significant in that it explores IR and IS in parallel, and it demonstrates how search tasks, as a concept, have been fundamental throughout each area of research. Furthermore, Vakkari uses research surrounding search tasks to bridge the gap between IR and IS, a need that has been expressed by many prominent researchers (Vakkari, 1999). This study surveys research and findings from a broad perspective, and identifies with studies that assimilate IS results with IR work.

Chapter 1 has presented the primary motivations for this study. The studies highlighted throughout this chapter embody our goal of investigating user interaction and system functionality from the perspective of domain-specific search tasks. Many advances in user-centered and task-oriented research, including the Watson and Vakkari studies, have been performed as part of text-based IR research. Video IR, on the other hand, needs to be explored independently from text and image IR works because users require additional support when retrieving digital video. This point is illustrated by Google, whose simplistic interface consists of one text box and one primary search button. An interface such as Google's isn't likely to be successful for this research because our focus is on supporting video search tasks from a particular professional domain.

An important underlying goal for this study, however, is to help promote design principles for supporting specific tasks and domains, in the hopes that research will eventually lead to the creation of interaction models. Exploring these problems is important because a preliminary survey has shown that no real difference exists across many video IR experiments and search systems, and results have been relatively unsatisfactory. The authors believe that insignificant findings and the lack of standards in video IR research are attributed to omitting contextual, i.e. domain or task-specific, parameters from systems evaluation and development. This research takes an exploratory approach for measuring the impact of tasks and domain on searching behaviors, user preference, and system performance and presents findings as implications for interface design and systems development.

Chapter 2

Problems

As previously discussed, one underlying goal of this study is to help promote the deployment of standards and principles in interface development, in the hopes that research will eventually lead to the acceptance of interaction models[14]. To achieve such results, it is necessary to begin evaluating video user interfaces from the perspective of domain-specific search tasks. The fundamental problems explored as part of this study reside on three basic levels. First, this study examines the problems and challenges associated with interface design principles for video retrieval systems. Next, problems related to domain-specific interface design principles are addressed. The authors then investigate interface design principles for K – 12 Science Education using a domain-centric approach. After presenting each of these problem areas, challenges facing

[14] Such models are needed to depict users in a video IR context, which refers to the searching and retrieval of digital video information by end users using interactive interfaces.

interface development, in general, are discussed. This chapter details the specific questions raised for this investigation.

2.1 Interface Design Principles for Video Retrieval Systems

There are multiple problems related to current principles for developing interface features and functions. Weaknesses in developmental efforts can be attributed to the lack of advancements in video IR experiments. Factors contributing to video IR's inactivity include: 1) a deficiency in powerful interface features and functions, and 2) few search experiments consider contextual information throughout system evaluation. To break this circular problem, researchers can begin analyzing video IR problems from the perspective of search tasks and examine user behaviors in different contexts. Once experimental results begin demonstrating distinctions across different tasks and individual interface features and functions, general design principles are then likely to emerge. Establishing interface design principles should be a top priority for video IR research so that future studies are given a practical foundation in which to build upon.

Several text-based IR studies have produced interface design standards, or applied frameworks. Examples of these frameworks can be found in studies by Shneiderman, Byrd, and Croft, 1997; Marchionini, 1995; (Toms, Freund, Kopak, & Bartlett, 2003), all of which are further discussed in Chapter 3. However, the scope of many of these studies is too narrow and not completely applicable for video IR research.

The limited number of user-centered video IR studies evokes additional problems for interface design. For example, a lack of fundamental user-centered principles in video IR causes researchers to rely on findings from text-based studies. Video IR needs to advance, and begin developing design principles based on video IR experiments and findings. Video IR principles are important because users require additional support, i.e. interface features and functions, when searching and browsing visual information (Marchionini & Geisler, 2002). Researchers can begin distancing video IR from textual IR by employing different strategies for designing and evaluating user interfaces.

Another challenge facing interface design principles is that most video IR studies are technologically driven. Many video IR studies attempt to advance content analysis applications where results are primarily measured in terms of systems performance, e.g. recall and precision. This characteristic of video IR research indicates that user-centered methods and principles are not a priority for many experimental studies.

Unfamiliarity with video search tasks is another factor that inhibits user-centered video IR research; many researchers believe that search tasks should drive the development of user interfaces and retrieval systems. The lack of knowledge about video search tasks hinders the development of high-level design principles. Task-specific features and functions should also reflect future users, or targeted audience.

One framework for interface design, which is particularly important for this study, formed the basis for the Físchlár Digital Video Library. Lee and Smeaton (2004) created an interface design framework by analyzing several high-level information seeking (IS) theories[15]. Through their analysis, Lee and Smeaton (2004) were able to establish the key features and functions for video retrieval interfaces. The features identified by Lee and

[15] This study is surveyed in the following chapter and the interface design framework is illustrated on page 20. Lee and Smeaton (2004) drives much of the discussion presented in Chapter 3.

Smeaton include: browsing and selecting a video, searching through a video, browsing within a video, video play back, and query reformulation (Lee & Smeaton, 2004).

Lee and Smeaton's (2004) research is significant because it analyzes video retrieval from a user-centered perspective. One limitation to these findings, however, is that their framework is primarily based on textual IR and IS[16] studies. Moreover, Lee and Smeaton's findings derived from IS models that were formulated by analyzing (user) interaction with text retrieval systems. Results of Lee and Smeaton's study may not be totally relevant for video IR research because their framework isn't based on visual searching.

This research study suggests that video IR is still too underdeveloped to begin formulating high-level interface design frameworks, similar to Lee and Smeaton (2004), or information seeking theory. Moreover, it is difficult to develop and justify general interface design principles because no high-level video IR theory exists. Further analyses into user interaction and search strategies are required before employing general interface design principles. Researchers can progress toward design principles by analyzing users throughout specific domains, or professional communities. Results from domain-specific studies will help develop user typologies and establish features and functions, or interface design frameworks, that are beneficial for specific user groups and applications. Future studies will then have a basis for forming high-level video information seeking models which, in turn, will support the development of general interface design principles.

[16] Information Seeking.

2.2 Domain-Specific Interface Design Principles

When developing domain-specific interface design principles, researchers must first understand what constitutes a domain. Examples of domains are quite broad and widespread. Essentially, a domain is a particular professional community or academic discipline. For example, waste disposal is an example of a professional domain while astronomy is considered an academic domain [17].

When designing domain-specific information systems, researchers and developers must understand that most domains will include a unique set of users. Rather than simply recognizing and analyzing different groups of users, researchers should also investigate the searching behaviors employed for finding domain-specific information. While both text and image IR researchers have analyzed information actions and search tasks across different domains[18], video IR is far behind because no significant task- or domain-centric study has been performed. Domain-specific video search tasks are wide-ranging and several examples are described in Chapter 1.

Why is it important to understand domain-specific tasks before designing interface features and functions? The authors suggest that retrieval systems designed to support domain-specific search tasks require more customized interfaces because users across domains will find different searching techniques important. For example, if someone was to develop an interface for a medical-research video database to be used by

---

[17] Many domains belong to both professional and academic communities.

[18] Text and image IR studies have explored information seeking throughout a variety of domains including general Web searching, American History, and Journalism.

practicing physicians, and another interface that allows crime investigators to search criminal records, the authors believe the interfaces developed to support each of these examples would vary considerably. Moreover, users from each domain will perform different tasks, thus require different interface features and functions. Criminal investigators may need a time-efficient, or fast, thumbnail browsing feature, while a medical researcher may need a more high-resolution, or slow motion, presentation of video information. These variances in information needs across different domains are likely to affect user interaction and should therefore influence the design of user interfaces.

Video IR research has not actively pursued domain-specific or task-related problems. Few, if any, studies have examined the implications of domain or task on systems design. The complexity of video information and video search tasks demonstrates the need for powerful interface features and functions. As stated in the previous chapter, a general interface, i.e. Google, isn't likely to be useful for this particular study because the emphasis is on supporting video search tasks and specific domains. Video IR researchers need to analyze search tasks and measure the effects on user behavior and search strategies. Findings from such studies should drive the creation of interface design principles.

2.3 Interface Design Principles for K – 12 Science Education

For the purposes of this study, the authors chose to investigate domain-centric problems within video IR by designing experiments around K – 12 Science Education search tasks.

Science education, as a domain, is particularly important for this study because it is generalizable and provides important challenges and services for the research community.

The authors assume that K – 12 science educators are interested in a broad array of topics and examine a wide variety of educational materials. Research associated with Science Education can span many different subjects from archeological discoveries to fitness in space. As you can imagine, K – 12 Science Education is highly interdisciplinary with relationships with chemistry, computer science, geography, physics, earth science, archeology, etc.

Researchers from the Open Video Project have already developed a system to retrieve video segments from the NASA K – 16 Science Education Programs. The Open Video Project previously employed graduate students to manually annotate and segment the NASA video. (Data representation and video structure are discussed in Chapter 3.) Although manual annotation and controlled vocabularies may be important for domain-specific video retrieval systems, the authors ask whether or not Science Education contains video needs and tasks that extend beyond keyword searching. This study also assumes that science educators have many reasons for retrieving video, thus a variety of interface features and functions may potentially benefit system functionality.

This study analyzes how K – 12 science educators search for and use digital video. A number of search experiments using the NASA K – 16 Science Education Programs, a public domain video dataset, and a system prototype were performed. To begin exploring interface design principles for Science Education, the authors first considered: who are the envisioned users of a Science Education video retrieval system, and what are their needs and tasks? Generally speaking, users of such a system will consist of science educators in practice, those currently in training (students), university faculty, and other educational researchers. As a result, a Science Education video retrieval system should be designed to support research and instruction-based tasks. By envisioning the tasks involved in Science Education instruction, systems developers can begin to formulate a variety of usage scenarios and information needs.

For the purposes of examining interface design principles for K – 12 Science Education, consider how a video digital library would benefit an upper undergraduate course from an education program. Essentially, one envisioned role for such a system could be to support a course titled, "Science in the Elementary Schools." Lessons throughout this course, including the development of student activities and teaching strategies, can be supported with digital video. The researchers of this study envision how instructors of a Science Education course could assign students to search for and analyze educational video, and create mock lesson plans and lab assignments which incorporate video content.

This example of a usage scenario depicts multiple user groups and domain-dependent tasks. This study measures how interface features and functions support similar users and tasks. Studies incorporating a user- and task-centered approach will be important because today's video IR studies are predominately technologically driven and fail to produce significant findings. Considering that this study broadens the scope of video IR evaluation and analyzes user interfaces based on the observations from specific domains and tasks, positive results were achieved. Once relationships between interface features and functions and domain-specific tasks become apparent, it then becomes possible to begin formulating interface design principles for Science Education. Findings from this study can support video annotation, indexing, interface development, and task analysis. (Specific interface features and functions that evaluated as part of this study are described throughout Chapter 4.)

2.4 Challenges for Interface Development of Video Systems

Designing user interfaces for video retrieval systems, in a general sense, presents many challenges. The challenges facing interface development for video retrieval extend beyond the problems facing text-based systems due to the complex nature of video information. Users can interact with digital video on a variety of different levels, such as retrieving and analyzing video using text, image, audio, moving pictures, and any combination of each.

One set of general challenges facing system development relates to video metadata, data representation, and content analysis, i.e. “back-end” tasks. The tasks for accumulating, organizing, and managing video metadata must be performed by system developers. Video IR researchers need to manage textual metadata, such as keywords, descriptions, and abstracts, along with a variety of content-based information.

Indexing textual metadata poses several unique challenges. First, system developers need to associate, or append, keywords with video content. Moreover, researchers need to use textual information to retrieve different segments of a video, including by shot, story, program, etc. (Description of video structure is described in Chapter 3.)

Visual, or content-based, information also requires a great deal of analysis. Content-based information can be automatically extracted from images and videos, where complex processing techniques are required to measure and compare visual qualities. Researchers must then determine which visual characteristics best support the envisioned users and tasks, and design system functionality accordingly.

Information organization and management must also meet the needs of the users. As a result, researchers need to segment video content appropriately for future use. For example, users may want to watch entire videos, lasting hours, or individual shots, lasting seconds. Alternatively, users may want to assemble a variety of clips from different sources to tell a certain story, answer specific questions, or receive informational reports. Preference for each of these retrieval features needs to be predetermined by system developers.

Challenges associated with displaying search results and video representation should also be addressed by researchers. Typically, researchers employ keyframes, i.e. thumbnail images, to represent a video clip. Using keyframes, however, requires system developers to evaluate different techniques for keyframe selection and extraction. For instance, researchers need to consider which keyframes best portray, or depict, video content. Some video IR experiments have previously used complex image processing techniques to automatically select keyframes. A simpler approach to automatically extracting keyframes included choosing the first, last, or middle frame from a video clip. Manual keyframe selection, on the other hand, is another common approach. Tradeoffs for each of these keyframe selection techniques include differences in time, effort, and accuracy of keyframe representation. Also, many video retrieval systems incorporate a variety of textual information along with search results. Similar to selecting accurate keyframes, system designers also need to determine which textual information helps users assess relevance of video clips. For example, retrieval systems can include controlled vocabulary terms, transcript information, abstracts and descriptions, categorical information, etc. within the video search results.

Each of these considerations for interface development, in the end, must be translated into usable features and functions. Search pages must be designed so users can effectively express queries and consume video content. To address these concerns, researchers can provide end-users with different levels of "query control," i.e. the choice between Boolean operators, search fields, etc. Video retrieval systems can also be designed to accommodate different techniques for playing video. For example, playing speeds, audio levels, video formats, streaming versus downloadable, and clip lengths are some of the playback features that can be implemented by researchers. While many of these questions remain unresolved, this study will attempt to provide additional insight into interface development for specific domains and search tasks.

2.5 Specific Problems and Questions for this Study

The problems and challenges associated with this study reside on multiple levels. First, this investigation explores users and tasks from a specific domain. Following, there is an analysis of the principles for designing interface features and functions, from both a general and domain-specific perspective. Next, this study examines issues surrounding video retrieval evaluation; specific experimental methods are fully described in Chapter 4.

As previously stated, the user group at the center of this study consists of future and current K – 12 science educators. For the purposes of this study, these users are assumed to demonstrate needs and tasks that are unique to Science Education and applicable in a general video retrieval context. Subsequently, the researchers of this study hope that science educators will show a variety of intended applications, uses, and allocations for digital video. The assessment of user demographics and more general characteristics were performed during the search experiments (described in Chapter 6).

Factors deriving from task characteristics motivate one challenge for this study, which is to explore: *what are the important tasks for K – 12 Science Education in a video retrieval context?* The authors hypothesized that these video retrieval tasks will be complex, wide ranging, and general or domain-specific. It is also assumed that these (Science Education) video search tasks extend beyond simply finding student activities and scientific documentaries, but also encompass information needs that range from specific to vague. For example, science educators may want to find a specific video title, or they may want to find simple up-close images of red soil. (A survey of visual search tasks is presented in Chapter 3.) Once these information needs, and motivations behind the needs, are fully understood, this study then analyzed the potential framework for Science Education video searching. This study also defines and categorizes other tasks related to Science Education video retrieval, including those that may include video use and manipulation.

User and task analyses make it possible to investigate a more centralized question of this study. Furthermore, this study then asks: *how do Science Education search tasks influence user interaction with video retrieval systems?* The impact of video search tasks on user interaction was investigated on two different levels, general video searching and domain-specific searching.

This study then measures how the characteristics of search tasks influence the use of specific interface features and functions. In addition, the interface features and functions that best support the observed search tasks are also investigated. As a result, the third question to this study includes: *how do users' search tasks translate into interface features and functions?* The authors explore how, based on search tasks, users gravitate towards certain interface features and functions. Associating search tasks with interface features and functions involves evaluating systems that employ both general video retrieval features and domain-specific features. A more direct question for this set of challenges includes: *what features and functions best support general video retrieval tasks, and what features and functions best support Science Education tasks?* These questions do not focus upon usability factors or web design issues, but are explored on a conceptual level.

In order to fully appreciate these challenges, let's consider the current state of video IR. Generally speaking, current interfaces for video retrieval systems implement browsing, playback, and visual and textual search features. Examples of these interface features and functions are illustrated in Image 1, a screenshot of ViewFinder. ViewFinder, a video retrieval system, was first implemented to search and browse general news video[19]. As illustrated, users are allowed to search using keywords and color information. The textual search feature was designed to retrieve by the video transcripts, where search terms are compared against a video's spoken words, and clips with matching keywords are returned. The color search uses keyframes to form content-based queries, and clips with similar color levels are returned[20]. This content-based search feature, as integrated into ViewFinder's Promote search, is referred to as a query by example (QBE). ViewFinder also enables hybrid retrieval, or searching by keyword and color information. The hybrid search feature, also integrated into the Promote search, includes a weighting mechanism where users emphasize importance between color and text. Users can also browse ViewFinder using several general categories, such as date, source, and the combination of date and source. However, whether searching or browsing ViewFinder for video, search results are returned by shot[21] and represented using keyframes[22].

ViewFinder demonstrates several key, or basic, components to video retrieval systems. When designing user interfaces, researchers must determine how each of these features should be implemented to support users and tasks. For instance, researchers must consider how much users will browse versus how much users will search. Video IR researchers must also determine when, in the search process, users are most likely to benefit from keyword searching versus content-based searching. It is also necessary to

[19] ViewFinder currently retrieves video deriving from ABC World New Tonight, CNN Headline News, and CSPAN programming.

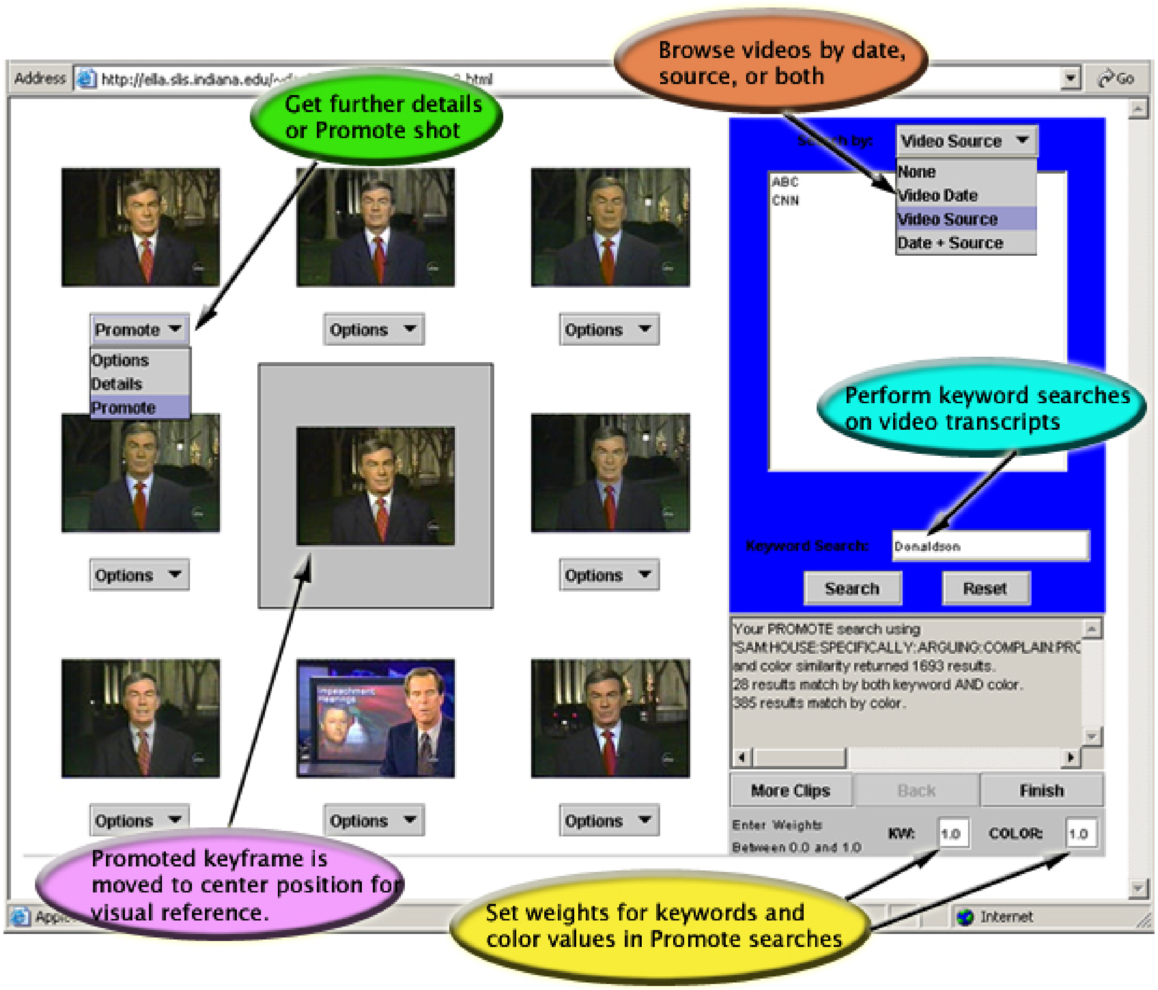
[20] The color search operates by image analysis and keyframe comparison.

[21] See section 3.3 for description of video document structure and a definition of "shot".

[22] Keyframes of ViewFinder were extracted as the middle frame of a video clip.

analyze which textual attributes are important for retrieval and which content-based features should be implemented. Researchers also need to investigate video representation and indexing. As described earlier, researchers need to determine what information, i.e. text, image, video, audio, etc., allows user to effectively assess relevance. This study does not, however, evaluate certain content analysis techniques, including keyframe extraction, video segmentation, and shot boundary determination, as each are considered beyond the scope of this study.

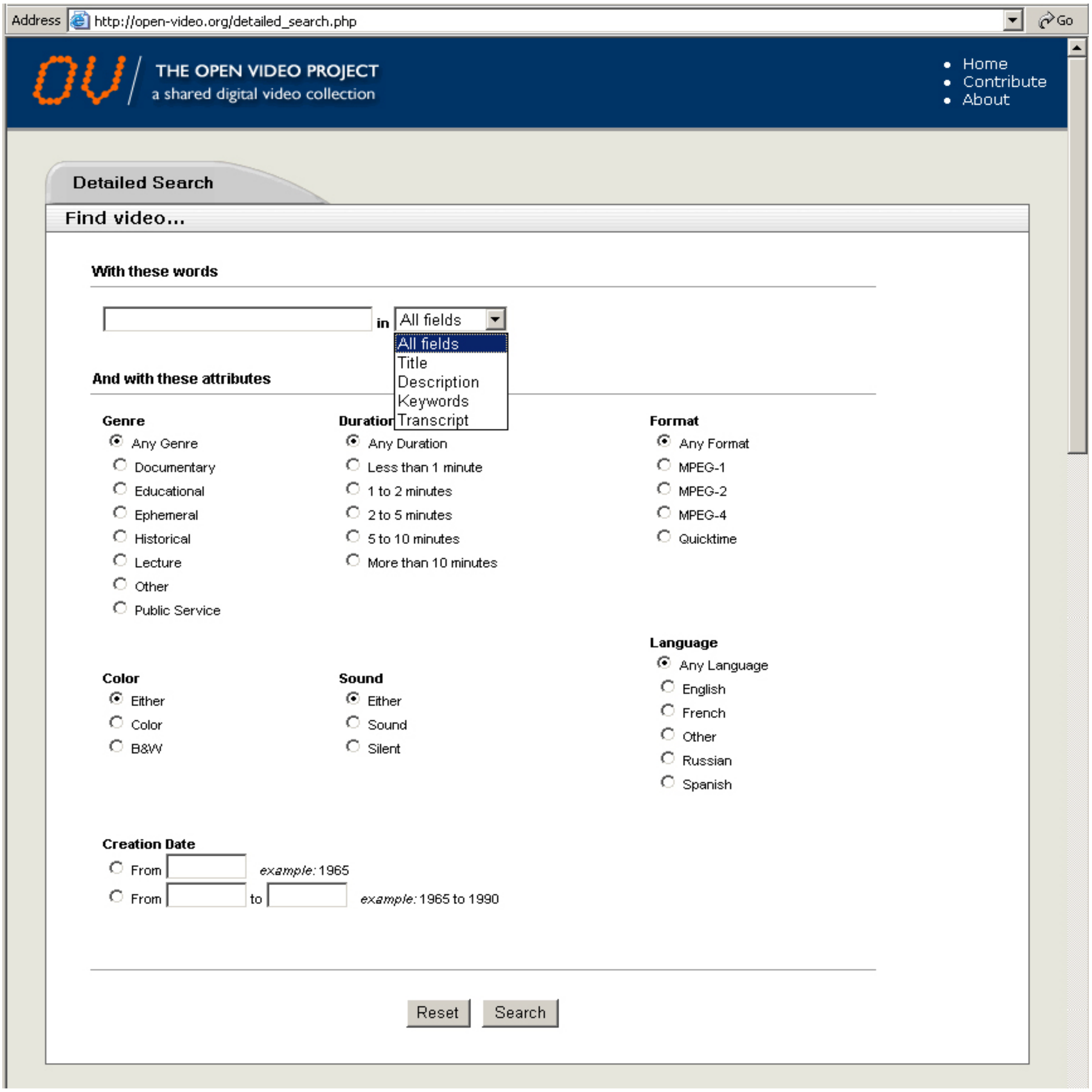
Image 1: Screen shot of ViewFinder, a news video retrieval system.



It is also important to understand the challenges facing domain-centered research. Open Video (http://www.open-video.org) researchers have developed a search system to retrieve video from the NASA K – 16 Science Education Programs. The current Open Video system retrieves video based on manual annotation and bibliographic indexing. Open Video's detailed search interface is presented in Image 2. From analyzing this system, readers can imply which features Open Video researchers deemed to be important for educational research. This study assumes that while the Open Video interface is practical for exploring task-centered video retrieval, there is a need for additional understanding of the features and functions that specifically support K – 12 Science Education.

Let's begin thinking about individual features and functions. From examining Image 2, it becomes apparent that Open Video researchers believe that keyword searching, based on manual annotations and bibliographic records, is essential for retrieving science educational video. Although, some concerns may arise when manually annotating digital video, this study assumes that Open Video's keyword search feature is beneficial as text-based search functions have been shown important for retrieving video. Manually annotation, although an arduous process, does provide high levels of information quality and analysis.

Image 2: Screenshot of Open Video's detailed search.



As previously stated, this study explores whether or not Science Education involves tasks that extend beyond controlled vocabularies, or keyword, searching[23].

[23] The experimental methodology, including data collection and analysis, along with a full description of experimental results are detailed in Chapters 4 and 6.

Moreover, some video search tasks may be better supported visually rather than textually. For instance, science educators may prefer searching visually, i.e. by color, shape, feature, etc., for vague video search tasks, or textually for specific search tasks. At this point, this study investigates which visual features are important for Science Education. A sub-question that arises from this analysis includes: for what types of search tasks are visual clues just as important as text? It will also be logical to ask: how, where and when is visual searching preferred over textual searching? Again, these questions were not developed to solely focus upon usability factors, but on a more conceptual level.

Several visual and textual features were subjected to evaluation. This study explored visual search features that retrieved video via color, shape, and texture. The textual components that were evaluated included a transcript-based search and textual query by example (QBE) feature. Combining these visual and textual search capabilities, a.k.a. hybrid retrieval, was also be evaluated. Prior to this study, the authors didn't know whether or not any of these techniques for searching and browsing science educational video would be effective, or if any visual features, specifically, would prove to be useful. However, based on the documented benefits of user- and task-centric methods in textual and image IR research, the authors believed that it would be important to conduct an exploratory video IR study from the perspective of users and search tasks (Budzik & Hammond, 2000).

The last set of challenges facing this study included systems evaluation. System performance measurements primarily focused on task performance and interface effectiveness. Searching experiments, described throughout Chapter 4, were designed to measure how effectively a system supports users and task completion. A specific question explored as part of this investigation includes: *how do the characteristics of search tasks affect or influence systems effectiveness?* This study also asks: *how do task-centered interface features and functions affect systems effectiveness?* Considering this study is being performed from a defined domain, a more logical question includes: *how do science educational search tasks and interface features and functions, designed specifically for Science Education, affect systems effectiveness?* The design of these experiments is important because domain-specific search tasks, a regularly overlooked variable, have been isolated and subjected to evaluation. Examining search tasks and domains as dependent variables in video IR studies may produce significant results for interface development. Correlations between user interactions, system performance and task characteristics makes this a significant video IR study whereas a vast majority of previous video IR research excludes these parameters from evaluation and does not subject them to comparison.

2.6 Problem Summary

This chapter has discussed the prominent problems and challenges facing this study. These problems reside on multiple levels. First, this study investigates principles for designing interface features and functions, from both a general and domain-specific perspective. Next, K – 12 Science Education, as a domain, is explored, and tasks related to Science Education are evaluated to determine how they can be employed for developing user interfaces. This chapter has also presented the specific questions raised by this study. The specific questions that addressed include:

- What are the important tasks for Science Education in a video retrieval context?
- How do the tasks identified for Science Education influence user interaction with video retrieval systems?
- How do users' search tasks translate into interface features and functions?
- What features and functions best support general video retrieval tasks, and what features and functions best support science educational tasks?
- How do search tasks and interface features and functions affect systems effectiveness, in a general sense and as they pertain to Science Education?

Chapter 3

Related Research

Prior to presenting the literature related to this study, readers should consider the complexity of investigating user interaction with video retrieval systems from the perspective of search tasks. Moreover, there are other research areas extending beyond those that have already been discussed that need to be explored to fully exhaust this investigation. For this particular research project, the authors needed to consider other studies including those that analyzed information seeking and general frameworks for designing user interfaces. Researchers also needed working knowledge of some of the technical components for developing video retrieval systems, including information management, client-server communication, and interface design tools.

As previously stated, the conceptual framework of this study explores how search tasks impact the use and design of interface features and functions which, in turn, may influence the creation of information seeking models. Each of these research areas are examined as they relate to video information retrieval. Readers should recognize that some of the problems surrounding this investigation have not been sufficiently explored in video IR research. The lack of previous research concerning video search tasks and video information seeking models causes researchers to rely on findings from text and image IR.

This chapter will survey various information seeking models stemming from text, image, and video retrieval research. Search tasks, a factor considered significant in most information seeking theory, will be described and analyzed throughout various domains. Approaches for developing interface features and functions of video retrieval systems will also be presented. Discussion of user interface research will coincide with technical aspects of system development and evaluation.

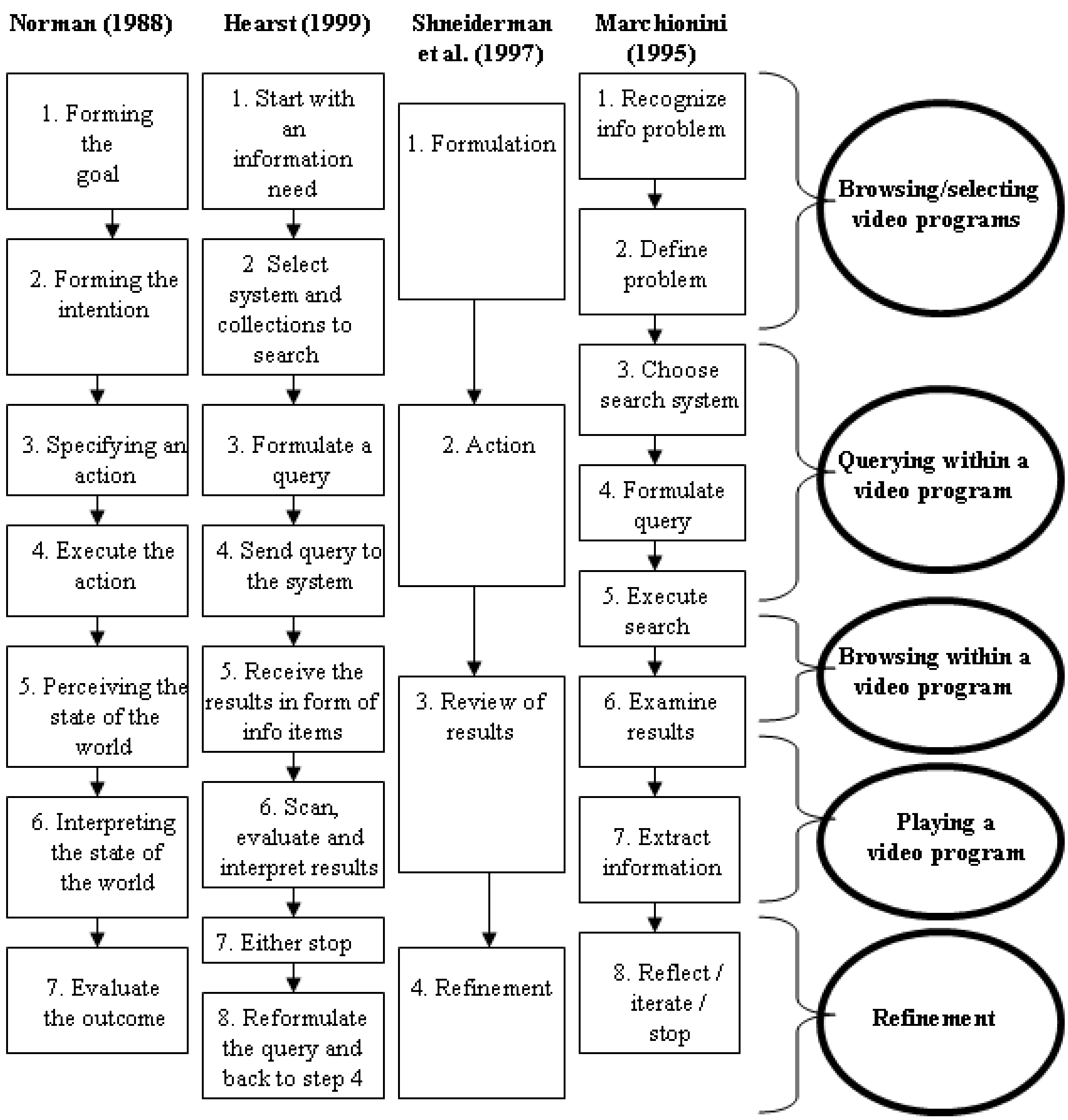
3.1 Information Seeking Models

3.1.1 Traditional Models

This section will discuss several theoretical models that have been examined and accepted by information science researchers. The models included in this analysis are shown in Figure 1. These models were analyzed in *Designing the User Interface for the Físchlár Digital Video Library,* a study which employed information seeking theory for developing an interface design framework (Lee & Smeaton, 2004).

Figure 1: Information seeking models and supporting interface functions.

[Source: Lee & Smeaton, 2004]



In their study, Lee and Smeaton associated a useful set of video search and browse functions with the stages of four behavioral models. The models examined by Lee and Smeaton included:

1) Norman's (1988) "Seven Stages of Action,"

2) Hearst's (1999) "eight sequences of interaction cycle"

3) Shneiderman et al.'s (1997) "four-phase search process"

Marchionini's (1995) "the various information-seeking sub processes"

Each of these models will be further discussed. Lee and Smeaton's study surveys one model, Norman (1988), which portrays everyday human actions, and three other models, Shneiderman, Hearst, and Marchionini, to further explain information seeking theory. None of these models specifically pertain to video searching. Due to the overlaps among text and video, and the lack of theory in video IR research, all of these models are considered important for this study. Moreover, each model is sufficiently abstract that applying them in video IR research is considered useful and appropriate. It is believed that these models provide a practical foundation for this study and each will contribute additional insight into user interaction.

3.1.1.1 Seven Stages of Action

Norman (1990) provides a framework for explaining how people perform everyday tasks. Norman's research, published in *The Design of Everyday Things*, doesn't specifically pertain to information seeking or human-computer interaction; however, his Seven Stages of Action serves as a good foundation for understanding user behavior and user-centered design.

Norman (1990) first discusses how mental models help humans explain certain phenomena, and how people formulate mental models based on prior knowledge and information. According to Norman, mental models form the basis for the Seven Stages of Action, and how humans approach everyday tasks (Norman, 1990).

The purpose of Norman's study is to form "an approximate model, not a complete psychological theory" (Norman, 1990). In recognizing this assumption, Norman identified several omissions from his framework including:

- the successful completion of all stages in sequential order is not required
- feedback throughout the process can result in "subgoals" or "subintentions"
- the process can begin at any point within the cycle and actions are frequently executed prior to the full development of goals

The Seven Stages of Action framework has been applied throughout a broad range of experimental studies, including those that explore user interaction, information seeking, and interface development. Norman's framework is relevant for interface development research, and this particular study, because it provides a greater understanding of how people perform certain tasks and what tools are most beneficial for users (Norman, 1990). Norman (1990) also claims that the Seven Stages of Action framework can be applied heuristically, or as a "basic checklist," throughout interface development. A heuristic developmental approach helps ensure that systems provide features and functions that best match users' goals and intentions. Researchers should note that employing Norman's theory for systems development isn't always a straightforward or perfect process as the assumptions surrounding his model, listed directly above, can significantly influence experiments and findings.

3.1.1.2 General Interaction Cycle

Unlike Norman, Hearst specifically analyzes information seeking behaviors. Hearst (1999) is, purposefully, abstract when describing and formulating a user interaction framework. In her study, information seeking is portrayed as a series of goals and tasks that include a wide range of activities and paths (Hearst, 1999). Moreover, Hearst simplifies information seeking and states that each model is basically an "interaction cycle consisting of query specification, receipt and examination of retrieval results, and either stopping or reformulating the query and repeating the process until a perfect result set is found" (Hearst, 1999).

Hearst (1999) then presents an information seeking model to reflect this assertion. Hearst's model, shown in Figure 1, is very general and takes into account several assumptions. The assumptions associated with her model include:

- environment being evaluated is Web-based
- user's information needs are always static
- information seeking process will continue to refine one query until perfect levels of recall and precision are achieved

These assumptions raise several questions and concerns. First, learning effect[24], which may cause users to skew "off-course," is not accounted for in this model. Also,

[24] Learning effect refers to users adapting to the retrieval system and/or gaining a better understanding of the information needs. Learning effect can alter the user's goals and intentions.

Hearst's model does not support browsing for "near-misses[25]" (Hearst, 1999). These omissions are significant because the concept of learning effect is widely documented and browsing for "near misses" is considered valuable for hyperlink-based systems, i.e. a web search engine (Hearst, 1999). In addition, Hearst's assumptions marginalize relevance feedback throughout the information seeking process. Bates (1989) and other information seeking researchers have challenged Hearst's model based on these assumptions.

Hearst's findings can still be useful when investigating information seeking and interface development. For example, emphasizing user's goals and task formulation can be attributed to Hearst. Hearst's model can provide additional insight into the features and functions that best support user's goals and tasks. Researchers applying Hearst's model throughout a systems design process should be aware of potential difficulties resulting from the assumptions listed above.

3.1.1.3 Four Phases of Searching

The primary motivation for Shneiderman's et al. (1997) information seeking research is to help facilitate the development of friendlier user interfaces. In his study, Shneiderman analyzed several popular text-based web search engines including AltaVista, Infoseek, Lycos, WebCrawler, and Open Text (Shneiderman, Byrd, & Croft, 1997).

---

[25] Near misses are relevant documents occurring in close proximity to the search results, i.e. connected with hyperlinks.

Recognizing the opportunity to advance interface functionality and design frameworks, Shneiderman and his team created a four-stage model, shown in Figure 1, which portrays user interaction with web search engines. Shneiderman's model was developed to help improve text-based search interfaces, and questions related to browsing functions were not thoroughly explored.

Shneiderman et al. (1997) described each stage of this model by presenting some traditional and futuristic interface features and functions that support information seeking theory. The interface features described by Shneiderman included a variety of technologies ranging from information visualization to search term truncation. Shneiderman's analysis serves as a good example of connecting information seeking theory with interface development. Readers are recommended to refer to the original source for further discussion (Shneiderman et al., 1997).

Another important finding from Shneiderman's et al. (1997) study includes certain recommendations, or "rules," for developing user interfaces. These rules form a useful framework for interface design. Shneiderman's rules call for:

1. consistency in design
2. short cuts for expert users
3. helpful feedback
4. exhibit searching closure
5. error handling
6. reversal of actions
7. give users sense of control
8. keep shot-term memory load low

It is possible for interface design researchers to apply Shneiderman's rules heuristically. These rules can help maintain standards in interface designs, which have been shown useful throughout experimental studies. Although Shneiderman's research is relevant for investigating interface designs, studies that evaluate both search and browse features may require additional support. In addition, researchers should keep in mind that Shneiderman's study only analyzed users in one particular context, i.e. using web-based textual search engines.

3.1.1.4 Sub-processes of Information Seeking

Marchionini (1995) is another well known researcher who signifies the importance of information seeking frameworks. In his book *Information Seeking in Electronic Environments,* Marchionini asserts that it is not feasible to explain each and every way people seek out information (Marchionini, 1995). Marchionini (1995) claims that in order to understand the diverse set of problems and actions related to information seeking, "it is useful to have a framework that explicates factors and processes common to information seeking in general."

Marchionini's analysis begins by reviewing some fundamentals of everyday life. Moreover, he discusses how we, as humans, build mental models to help explain the inner workings of our environment, and that our mental models are formed by information. Thus, Marchionini claims, "information seeking is a process driven by life itself."

It is our cognitive makeup that should guide the development of information systems (Marchionini, 1995). Moreover, user interfaces should be designed to support our instinctive way of seeking information, not according to available technologies. Findings taken from Marchionini, and other researchers such as Dervin (1977), Belkin (1980), and Kuhlthau (1988), contribute to this objective.

Next, Marchionini describes several factors associated with the information seeking process. These factors include “the seeker, task, system, domain setting, and outcomes” (Marchionini, 1995). The impact of each factor on the search process is discussed. It is important to remember that these factors are not isolated from one another and many complex relationships exist between them (Marchionini, 1995).

Marchionini’s information seeking model, shown in Figure 1, is considered to be “systematic and opportunistic.” Similar to the other models presented in this chapter, Marchionini’s model is made up of sub-processes. In addition, each model described in this section is initiated with an information need, and progresses until the need is satisfied or the process is discontinued. Marchionini’s model allows the sub-processes to be performed in sequence, sporadically, or in parallel. Marchionini (1995) describes each sub-process and demonstrates the dependencies, effects, and influences surrounding them. He then classifies the sub-processes into three categories including “understanding, planning and execution, and evaluation and use” (Marchionini, 1995). The reader is recommended to refer to the original source for further reading.

3.1.2 Visual Models

Visual models are designed to depict both image and video information seeking behaviors. Image and video information seeking models are germane due to the overlaps among data characteristics and retrieval systems. As a result, exploring image information seeking models is important for this study. A survey of both image and video information seeking models follow.

A significant amount of video IR research has emerged from the IBM Almaden Research Center. In a previous study, Amir et al. (2003) presented IBM's user interaction framework for developing video retrieval systems. Amir et al. (2003) claimed that video searching worked on two levels: visual searching and keyword/transcript searching. Although many IR researchers use generic frameworks to guide systems development, Amir asserts there are no collective frameworks for explaining video system interaction. Amir et al. (2003) propose a "generic paradigm" for developing retrieval systems that suggests video transcripts work best for searching, visual information works best for browsing, and a combination of each has the greatest potential for success.

Amir's developmental framework takes into account a set of simple information seeking activities, or user actions. These information seeking activities include: query construction, scanning the search results, and examining any selected documents (Amir, Srinivasan, & Efrat, 2003). Amir et al. (2003) assert that these activities become much more complex when searching visual information. The complexity of visual searching comes from users having to comprehend more information throughout the search process.

Moreover, users can consider text, image, video, and audio information when assessing relevance of video search results.

Simonnot and Smaï (1995) also provided a framework to depict visual information retrieval. Their framework represents an interactive, i.e. "human in the loop", process for retrieving satisfactory visual information. Simonnot and Smaï (1995) state that traditional retrieval models are often recycled for a variety of research areas and systems, regardless of their focus. This characteristic of visual IR research is not ideal because text, image, and video retrieval systems should incorporate different search functions. For example, relevance feedback features have been shown to be powerful for image search systems, and user interfaces should be developed to reflect such findings (Simonnot & Smail, 95).

Simonnot and Smaï (1995) recognize that because of the complexity of managing and indexing video data user's interactivity and search strategies are more important for retrieving video than for text. Moreover, it is very difficult to implement video retrieval systems that automatically[26] return high-quality search results. Therefore, having models that depict video information seeking or interaction becomes important. Simonnot and Smaï's video information retrieval model, presented in Figure 2, is distinguished from other models because it associates system functions with user actions and is supported by relevance feedback.

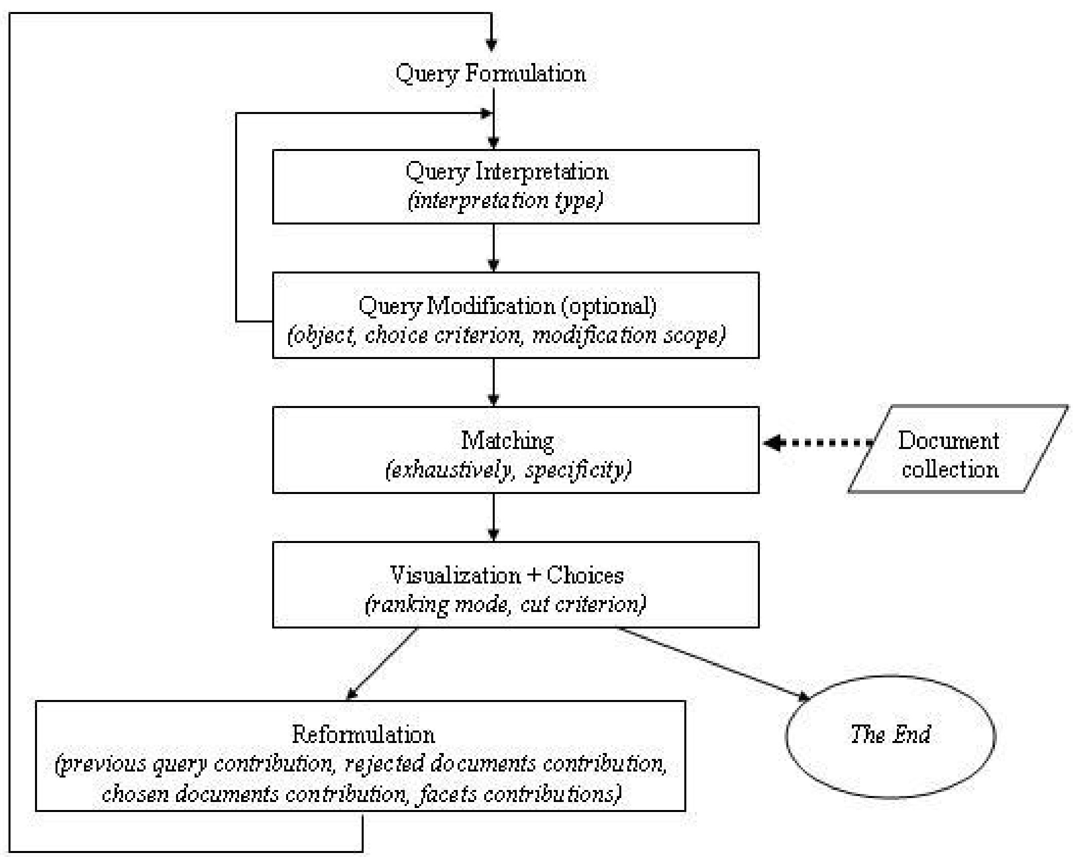
One benefit from Simonnot and Smaï's (1995) study comes from emphasizing relevance feedback for video retrieval. Relevance feedback helps bridge the "semantic gap" between human perceptions and the content-based qualities of images and videos.

[26] "Automatically" meaning without feedback from the user or requiring high levels of interactivity.

The semantic gap is a “misunderstanding” that occurs when low-level information[27] is used to retrieve semantic[28] concepts and vice versa. The semantic gap has been recognized and studied by a number of researchers including Amir et al. (2003).

Figure 2: Video information retrieval process.

[Source: Simonnot & Smaï, 1995]



Conniss, Ashord, and Graham (2000) conducted the most comprehensive visual information seeking study to date. Their experiments and findings, published in

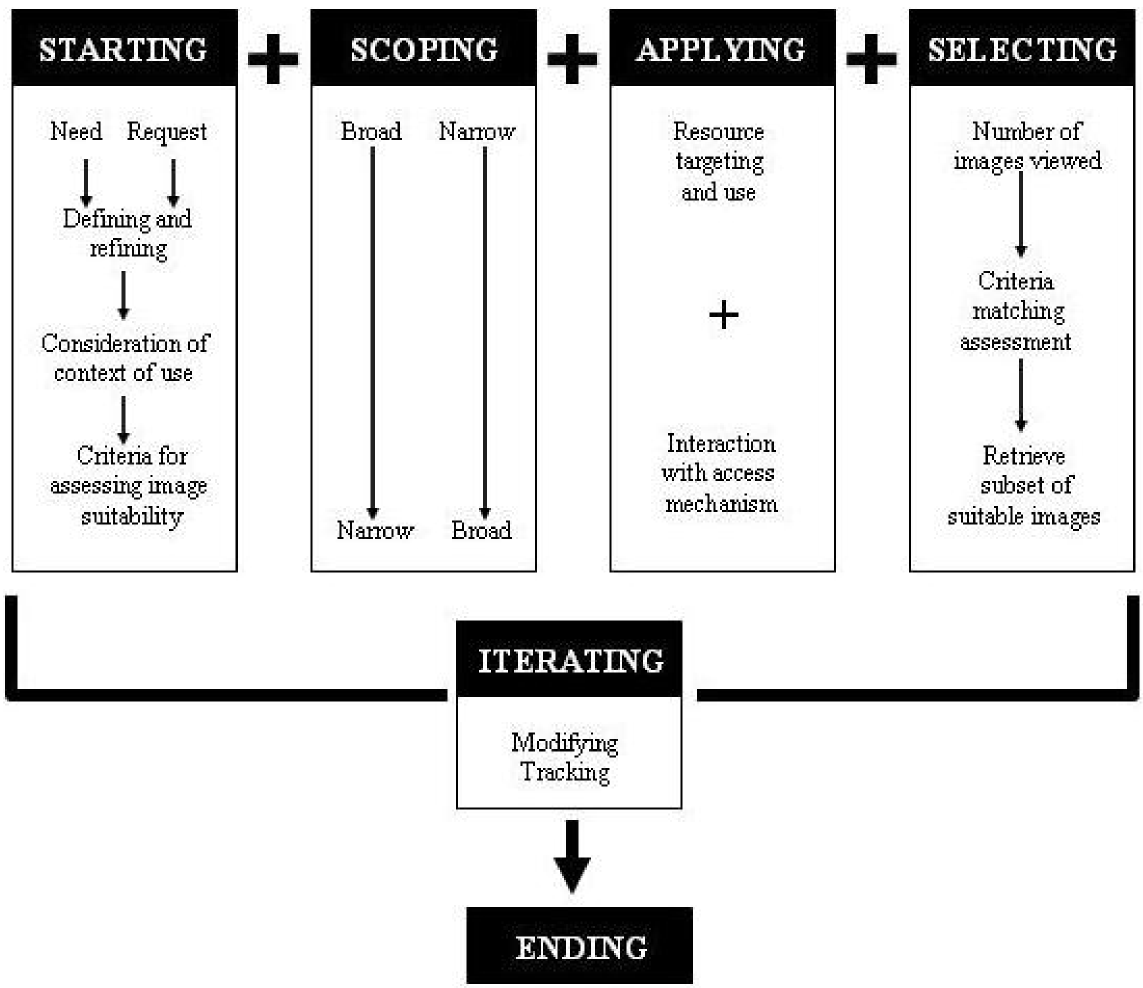
[27] Low-level information usually refers to content-based information, i.e. colors, shapes, features, textures, etc.

[28] Semantic information for video retrieval is typically found in textual information including manual annotation and video transcripts.

Information Seeking Behaviour in Image Retrieval: VISOR I Final Report, provide thorough analysis of why and how certain individuals search for and use image information. Conniss et al. (2000) explore problems related to image seeking across several domains that were recognized to have notable influence on user behaviors.

Figure 3: Image information seeking model.

[Source: Conniss et al., 2000]



Conniss et al. (2000) present a general framework, shown in Figure 3, to describe image information seeking. The information seeking behaviors associated with this framework are highly dependant on a number of different factors. These user-centered factors include: workplace, existing resources, personal experiences and influences, communications and interactions, and access mechanisms (Conniss, Ashford, & Graham, 2000). A number of complex relationships exist between these factors and the reader is encouraged to refer to the original source for further details.

The primary phases of Conniss's framework include: "starting, scoping, applying, selecting, iterating, and ending" (Conniss et al., 2000). These phases are not consecutive or distinct, but form an ever-changing and cyclical process (Conniss et al., 2000). Each phase is made up of a series of sub-processes and decisions. The sub-processes also inherit a set of complex relationships where every decision can influence the succession and operation of the information seeking process.

Conniss's study is also helpful for video retrieval research. General frameworks, such as Conniss's, provide valuable insight into interaction behaviors and decision making processes. As a result, information seeking frameworks can support the development of user-centered systems and, subsequently, transparent interface features and functions. Findings from this line of theoretical research will also support the development of standards for developing visual retrieval interfaces. Standard interface features and functions will be discussed later in this chapter.

3.2 Search Tasks

Investigating search tasks, as they relate to user interaction and interface design, is challenging work. A lot of IR researchers have explored task-related problems by analyzing individual information need statements and search queries. While analyzing information needs provide insight into system use and relevance assessments, researching search tasks is more complex and requires further analysis. Researchers need to consider that search tasks are multidimensional and encompass only one element of the users' context. Consequently, other factors surrounding users' context, including domain, goals, and application(s), may also need justification in search task-related research. Studies that focus on video search tasks need to acknowledge the complexity of video information. Moreover, researchers should consider how the composition and structure of video affects search tasks and interaction behaviors, and vice versa.

This survey will cover text-, image-, and video-related studies. Text and image IR research is also important due to the overlaps among data characteristics and the lack of progress in user-centered video IR research. Image IR research is particularly relevant for this study because several experiments have examined search tasks from a visual perspective.

3.2.1 Text Searching

A number of text-based studies have examined the influence of search tasks. Budzik and Hammond (2001), discussed in Chapter 1, suggested that users' context can provide valuable information for retrieving documents. To test this hypothesis, Budzik and Hammond (2001) implemented Watson, an information management assistant. The Watson system analyzed user's search tasks and interaction behaviors, and intuitively retrieved documents. Watson was integrated into a word-processing application and several experiments were performed (Budzik & Hammond, 2000). Results showed that Watson's approach for analyzing and applying user's search tasks for retrieval was beneficial. (Refer to Chapter 1 or original source for further details about the Watson experiments.)

Influences of searching domain have also been explored throughout textual IR research. Toms, Freund, Kopak, and Bartlett (2003) investigated searching behaviors across several popular Web domains. The authors assumed that understanding search tasks, strategies, and interaction behaviors was important for developing supportive user interfaces (Toms et al., 2003). Toms et al. (2003) anticipated that their study would help establish a set of task-oriented Web search features.

Toms et al. (2003) designed user-centered experiments to evaluate the four most popular search domains, including consumer health, shopping, general research, and travel. Forty-eight participants were presented four search topics[29] from each of the four domains, for a total of sixteen topics per participant. While performing the search topics, each participant was given the option to choose between Google's keyword search and categorical browse features. Results were measured using qualitative and quantitative methods. Pre- and post-test questionnaires, audio-taped interviews, server logs, screen captures, and verbal protocols were all used to collect experimental data.

Toms's et al. (2003) results showed significant differences among user interactions across each of the search domains. Users spent more time perusing individual websites when searching for shopping and travel information. General

[29] The search topics originated from the 10th annual Text REtrieval Conference's (TREC 10) Interactive Track (Toms et al., 2003).

research and consumer health topics indicated that users preferred to keyword search and, subsequently, browse search results[30]. Results also exhibited that users favored browsing categorically for shopping and travel topics (Toms et al., 2003). These findings demonstrate that searching behaviors vary across different domains, and that one simplistic interface, i.e. Google's keyword search, isn't always the most practical for all search tasks.

High-level research involving users' search tasks dates back several decades. Vakkari (1999), as described throughout Chapter 1, comprehensively surveyed search tasks throughout IR and information seeking (IS) research. Vakkari (1999) analyzed important IR studies including: Bates, 1989; Belkin, Seeger, & Wersig, 1983; Belkin, 1980); (Saracevic, 1996. Important IS studies surveyed by Vakkari included: Dervin & Nilan, 1986; Kuhlthau, 1993; Kunz, Rittel, & Schwuchow. Vakkari (1999) used task complexity to draw associations between these two traditionally separate[31] fields.

Vakkari (1999) analyzed tasks from the perspective of problem solving. Search tasks were recognized as having subtasks, and being components of larger jobs. Search tasks exhibited finite starting and ending points, and contained a variety of goals and actions. While performing search tasks, users assess information needs, relevancy, and search strategies, in other words, carry out information actions (Vakkari, 1999).

Vakkari (1999) claims that the information actions – as just listed – are all related to task complexity. Moreover, as task complexity increases, pre-determinability of information needs, searching process, and satisfying results decrease, and vice versa.

[30] These results were not statistically significant

[31] Separate according to task-oriented, process-oriented, and document-oriented studies and/or systems.

Task complexity is based on task repetition, analyzability, available paths, and outcomes. In addition, user's perception about the pre-determinability, or uncertainty, of search tasks can portray task complexity (Vakkari, 1999).

Vakkari (1999) also argues that prior knowledge and problem structure are also related to task complexity. Task complexities, as examined by Vakkari, ranged from the most difficult to the easiest tasks, including genuine decision task, known-genuine decision tasks, normal decision task, normal information processing task, and automatic information processing (Vakkari, 1999). Researchers can deduct that prior knowledge provides insight into problem structure, information needs, and information actions. Problem structure can also convey prior knowledge about search tasks and indicate search strategy.

3.2.2 Visual Searching

Image IR research has made significant progress investigating visual search tasks. Moreover, image IR researchers have tried to understand searching behaviors by exploring different visual information needs.

Several researchers have examined visual search tasks by identifying and classifying information needs. Classifying visual information needs, or search queries, has been performed for a variety of reasons, most notably to improve retrieval algorithms or indexing techniques (Jörgensen, 1996) (Yan et al., 2004); (Yang, Chaisorn, Zhao, Neo, & Chua, 2003).

Studies have shown that visual information needs range from very general to very specific. General[32], or abstract, information needs are conceptually vague[33] and typically require additional input from the user to estimate relevance. One example of a general information need includes, "find images that symbolize jealousy." For such a need, users' interpretation of search results would be essential for selecting relevant images.

Other information needs are less vague and typically have a predetermined, or common, criteria for assessing relevance. A specific image need could be "find image 40 in gallery 66 of collection 14A1," while a specific video need may be "retrieve shot 18 from video 44" (description of video structure will be discussed later in the chapter). These examples contain little ambiguity and provide users and evaluators clear ideas of what information is relevant and what is not.

The general and specific information needs described in the previous paragraph are extreme examples. Most other visual needs fall somewhere in between. A framework for visual information needs, as identified throughout video and image IR research, is presented in Figure 4. As shown in this diagram, there are different levels of general and specific information needs. General information can depict a particular subject, i.e. globalization, or contain certain visual characteristics, i.e. red circles. Specific information can include images or videos that show a particular object or event taking place. For example, users may want a video clip of the Statue of Liberty at night, or the opening kickoff at Super Bowl 40.

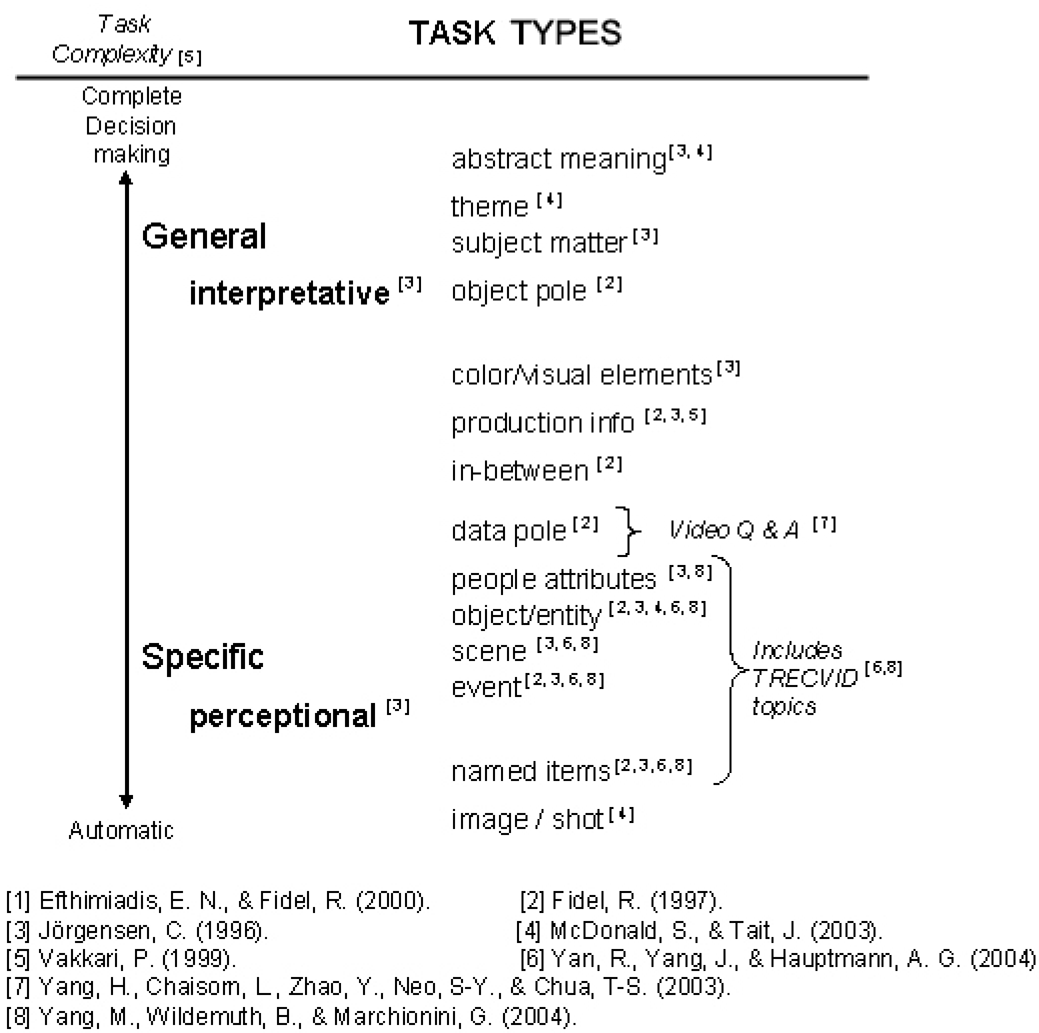
---

[32] Differences in terminology exist across visual and textual IR research.

[33] A conceptually vague information need could be a request for images that depict an abstract meaning or theme.

Task-focused image IR research has progressed beyond identifying and classifying information needs and search queries. Several studies have connected search tasks with impacts on user interaction, search satisfaction, and other measured variables.

Figure 4: Framework of image and video search tasks.



McDonald and Tait (2003) investigated how users formulated color-based queries for different image search tasks. The influence of search tasks on system use and search performance[34] was measured across two experimental sessions. Three types of search tasks, including previously seen images, specific search, and general search, were created to explore these questions. McDonald and Tait (2003) analyzed whether or not these particular search tasks would influence users to select a sketch search tool, a color browse, or serial browse feature (discussion of interface features will be covered in the next section).

McDonald and Tait (2003) discovered a significant association between search task and tool selection. For the previously seen search task, users employed the color-browse and sketch features equally. Results also showed that users preferred the serial browse feature for specific tasks and the color browse for general tasks. These results also hold true for correlation between interface feature, i.e. search tool, and search success[35] (McDonald & Tait, 2003). Search tasks were also found to influence search success (McDonald & Tait, 2003). Previously seen search tasks led to the highest number of successful searches. General tasks showed the second highest levels of search success and, to the surprise of the authors, specific search tasks led to the least amount of successful searches. Readers can refer to the original source for further details.

Yang, Wildemuth, and Marchionini, (2004) explored similar problems using the TRECVID evaluation protocol. Yang et al. (2004) evaluated how search topics[36]

---

[34] Search performance in McDonald and Tait (2003) was measured according to user's criteria not systems performance.

[35] Browsing was found best for specific and general search tasks and there was no difference between color-browse and sketch for "previously" seen image task.

[36] TRECVID search topics refer to common information need statements.

influenced the performance[37] of three video retrieval systems. Their experiments consisted of having 36 participants perform a total 12 TRECVID search topics across a hybrid, transcript-based, and visual-only search system, i.e. 4 topics on each system.

Yang et al. (2004) classified each of the 24 TRECVID search topics as being either generic or specific. Generic topics sought general, or unnamed, phenomena, such as “roads with vehicles” (Yang et al., 2004). Specific topics, on the other hand, included named items, locations, or persons, including “Pope John Paul II” or “The Sphinx”.

As discussed in Chapter 1, Yang’s et al. hypothesis that the hybrid system would significantly outperform the transcript and visual-only search systems was not supported. Overall, all three systems performed similarly for generic topics; the only significant difference was observed between the hybrid and visual-only search systems (Yang et al., 2004). For specific topics, the transcript-based system outperformed the other two systems, and the visual-only search system was again least effective[38].

Yang’s et al. results were not all that revealing because no real difference between search systems was observed, and text features were shown to be the key element for retrieving video (Yang et al., 2004). In addition, Yang’s et al. study was restricted to the TRECVID evaluation protocol where, when compared to other task-focused research, all search topics are quite specific. Moreover, it is still too difficult to commonly evaluate

---

[37] System performance in Yang et al. (2004) was measured according to TRECVID evaluation, i.e. various levels of recall and precision.

[38] The retrieval performance for the transcript-based and hybrid search systems significantly improved for specific search topics. Performance of the visual-only search system remained relatively the same for both types of search topics.

video retrieval using general, or conceptually vague, search tasks and TRECVID doesn't support such experiments.

Yang's et al. study is still important however because it is one of few video IR experiments that attempts to look at problems from the perspective of search tasks. CMU and the National University of Singapore (NUS) performed similar experimental studies, discussed in Chapter 1, which also analyzed TRECVID search topics. The purpose of the CMU and NUS studies was to measure systems-centered deliverables including machine learning and video retrieval algorithms. Few video IR studies explore search tasks found throughout specific domains, or attempt to enhance user-centered research.

Image IR researchers have investigated search tasks throughout the fields of journalism, advertising, and American History. Ornager (1995) used a Danish newspaper database to analyze the visual queries of journalists and the cataloging protocol for news images. Ornager's study examined 13 newspaper archives, 25 archivists, and 26 journalists, and employed both quantitative and qualitative methods[39]. Results of this study supported various categories for image archiving and helped develop a user typology for journalists. Findings from Ornager (1995) can benefit image indexing and interface development for news retrieval systems.

Garber and Grunes (1992) studied how art directors of advertising agencies searched for images. Garber and Grunes (1992) conducted a work flow and task analysis of art directors and – based on their findings – formulated a visual search model. Next, they developed a user interface and presented it to a series of focus groups. The focus groups showed the prototype interface was promising and held potential to streamline

---

[39] Methods included interviews and observation.

visual searching for art directors (Garber & Grunes, 1992). Garber and Grunes (1992) also employed a variety of interviewing and observation methods throughout their experiment.

Choi and Rasmussen (2003) analyzed the visual search tasks of American History scholars[40]. This study helped identify image attributes that could be useful for indexing and retrieval. Results indicated that users employ more query terms when searching for images than when performing everyday Web searching. Results also showed that historians widely used search terms related to image subject. Image searching based on creator, author, and title was not considered important; however, historians did prefer to search by place, event, condition, location, and time (Choi & Rasmussen, 2003).

An important video IR study, beyond the scope of TRECVID, is beginning to investigate video retrieval for ethnomusicology research. The EVIA project at Indiana University is digitizing and manually annotating raw, i.e. field, footage of musical performances from all over the world (EVIA Digital Archive - Ethnomusicological Video for Instruction and Analysis, 2005). EVIA is important because it is one of the first of its kind where researchers are beginning to investigate video retrieval within a specific domain. The EVIA system will be based on annotations provided by ethnomusicologists and will support search tasks of faculty, students, and other researchers (discussion of video retrieval systems will follow). The EVIA project should have broad implications for user modeling, metadata creation and indexing, and experimental designs, which can also be employed for other interrelated projects.

[40] Choi and Rasmussen (2003) evaluated the queries of 38 faculty and graduate students.

3.3 Interface Features and Functions

Now that this paper has presented research from a user-centered perspective, it is also important to present research pertaining to systems. There have been many interface features and functions evaluated throughout video IR research. However, investigators need additional understanding of how interface features and functions of video systems support specific tasks and domains. For example, users searching for video using face recognition information would likely employ different search strategies than someone searching for sports highlights. Researchers from different domains need to understand the interface features that are available and preferred by users. This section will review the interface features and functions implemented for a variety of video retrieval systems. It is also inherent that this chapter discuss issues surrounding system functionality and information management.

The interface features and functions described throughout this section were implemented by regarded researchers and laboratories from both industry and academia. Many of the projects highlighted in this survey participated in the TREC Video REtrieval Evaluation Workshop, known as TRECVID. Systems from Carnegie Mellon (CMU), Dublin City University (DCU), IBM, the University of North Carolina at Chapel Hill (UNC), and others are surveyed in this section.

Generally speaking, several principles are typically strived for when developing user interfaces. Conniss et al. (2003) express the need for flexibility[41], which can help reduce a system's learning curve. For example, interfaces should offer the speed and

[41] Flexibility refers to capability of supporting a wide range of uses, needs, environments, and user groups

power desired by expert users while providing the straightforwardness and ease-of-use expected by system novices. Questions such as these must be considered when designing flexible interface features and functions.

For the purposes of this survey, we will equate flexibility with generality. Marchionini and Geisler (2002) identified three general levels of interaction with video retrieval systems. These levels include searching, browsing, and contributing; each can be translated into general functions for user interfaces (Marchionini & Geisler, 2002).

Video search functions come in two different varieties, including text and content-based. Text-based video searching can be implemented using multiple features and functions. One example of a text-based video search function includes a typical keyword search[42]. Keyword search functions are common and have proven to be effective when applied to video transcripts, or automatic speech recognition (ASR) and closed caption (CC) outputs. For example, DCU, IBM, CMU, and UNC incorporated a transcript-based search feature into their TRECVID 2004 systems. The search systems corresponding to each of these teams include:

- Físchlár Digital Video Library (DCU)
- CueVideo (IBM)
- Informedia (CMU)
- The Open Video Project (UNC)

Textual searching for video systems is not limited to transcripts. Some systems allow users to search other descriptive fields, or video attributes. For example, the Open Video Project (UNC) contains a “detailed search” where users can query video

[42] From the interface perspective, users formulate a keyword query and execute the search.

descriptions, abstracts, or titles. Users can also narrow search results by specifying genre, producer(s), and duration (Marchionini & Geisler, 2002). Lee and Smeaton (2004) comprehensively surveyed existing video retrieval systems and identified others that contain similar features. Systems identified by Lee and Smeaton include: VideoStar (Hjelsvold, Lagorgen, Midtstraum, & Sandsta, 1995), WebSEEK (WebSEEK: Content-Based Image and Video Search and Catalog Tool for the Web, 2004), and Internet CNN Newsroom (Lee & Smeaton, 2004).

Content-based video retrieval has also been the focus of much investigation. As previously described, content-based retrieval uses the visual characteristics of images or videos to retrieve information. The most common content-based retrieval techniques include searching by color, texture, edge, shape, and feature. The Físchlár system (DCU) previously implemented three variations of a color histogram search and an edge detection search (Browne et al., 2004). IBM also experimented with content-based retrieval by incorporating a shape, color, texture, and edge search into their TRECVID 2003 system (Amir et al., 2004). CMU's Informedia system included different variations of a color, texture, and edge direction search (Hauptman et al., 2004). Although there are similarities among the content-based retrieval features of these systems, each group employed different design and implementation strategies (the reader is recommended to refer to the original sources for details).

Video retrieval systems also support browsing. Browsing features can be beneficial for exploring a video collection or examining search results. Exploratory browsing can be facilitated with video transcripts, i.e. using the significant spoken keywords, manually annotated data, and visual information. Open Video allows users to browse the video collection by keyword, date, genre, actor, and contributing organization (Marchionini & Geisler, 2002). The Informedia and WebSEEK projects also provide similar exploratory browsing features (Lee & Smeaton, 2004). McDonald and Tait (2003), on the other hand, evaluated a system that enabled a color browse feature.

Browsing features also allow users to review search results. Retrieval systems use a variety of techniques to present video information. Search results can be tailored to reflect the various levels of video, including shots, scenes, video documents, etc.[43] Figure 5 illustrates how a video is structured and makes apparent that:

- entire video documents are made up of a series of scenes and shots
- scenes are made up of shots that are related semantically[44]
- shots are made up of a number of frames for one pan or focus of the camera[45]

Many video retrieval systems assign individual keyframes to represent the different levels of video information, or search results. Keyframes are still images extracted from videos. Displaying keyframes can be beneficial for users as they are able to visually scan and compare the contents of videos. Technically, there is no difference between keyframes and other frames; however, keyframes are selected because they have been deemed to contain visual characteristics or semantic value that accurately depicts a shot, scene, etc.

---

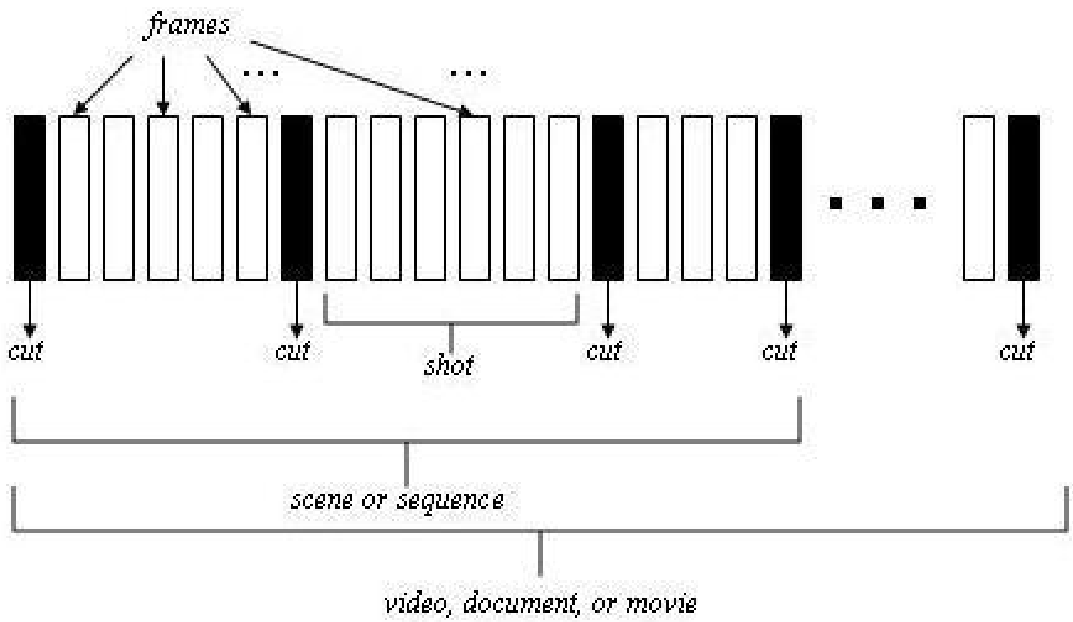
[43] A difference in terminology exists among studies, other concepts include: "action," "story," "event," "tape," etc.

[44] Shots comprising a scene or story do not have to be contiguous.

[45] Figure 5 only illustrates shots that begin and end with hard cuts. Shot transition can also include camera fades or zooms.

Figure 5: Video document structure.

[Source: Revised from Simonnot & Smaï, 1995]



There are a variety of ways video retrieval systems employ keyframes for browsing. One common visual browsing technique is a storyboard, implemented for Open Video, Informedia, and CueVideo. A storyboard displays multiple, semantically related, keyframes within the same results or browsing page. Another popular browsing technique is called a slide show. A slide show displays one keyframe at a time, which continuously changes after short intervals of time[46]. The Físchlár, Open Video, and CueVideo systems have all implemented and evaluated a slide show feature. Fast forwards is another common browsing feature. A fast forward feature, which has been implemented for Open Video and CueVideo, performs as its name implies: video clips are played back at a faster than normal rate to provide users a sense of motion and narrative (Marchionini & Geisler, 2002).

[46] Some slide shows are also accompanied with audio.

The Físchlár system includes a couple of unique browsing features. These browsing features include a timeline and a dynamic overview browser (Lee & Smeaton, 2004). The timeline browser allows users to select a set of keyframes, or clips, which match specific time frames of a video. The dynamic overview browser works similarly to that of a slide show; however, multiple images are presented within the display panel and when the user mouses over a particular keyframe the corresponding slide show is set in motion (Lee & Smeaton, 2004).

3.3.1 Query Models

There are many features and functions that facilitate video searching. In addition, users need to be able to express queries using different video attributes. Approaches for textual and visual query modeling will be the focus of this section.

3.3.1.1 Textual Query Modeling

Features developed for textual searching are considered important for video retrieval systems. For example, some keyword search features have enabled video seekers to employ techniques such as natural language and controlled vocabularies to express queries. Lee and Smeaton (2004) identified a number of video retrieval systems that contained a natural language search including Informedia, WebSEEK, Internet CNN Newsroom, Video STAR, and VideoQ.

The Open Video currently contains a controlled vocabulary search that performs a logical OR search by default. Both the Físchlár and IBM TRECVID system offer a full range of Boolean search capabilities including the use of AND, OR, and NOT operators (Adams et al., 2003). A modification to Boolean search capabilities was implemented by Simonnot and Smaï (1995). Simonnot and Smaï (1995) allowed users to express Boolean logic by weighting various elements of a search query including the various terms, facets, and categories. Query weighing was based on a -1 to 1 scale where 0 conveyed neutral, less than 0 conveyed undesirable, and greater than 0 conveyed important (Simonnot & Smail, 95).

Researchers from IBM evaluated several other query refinement and expansion features. Moreover, an exact phrase matching feature has been implemented for versions of the CueVideo system (Amir et al., 2003). CueVideo was also used to test three query expansion features including a thesaurus, cluster labeling, and a phonetic sound generator. These features provided further connections between user's queries and video information.

3.3.1.2 Visual Query Modeling

Visual searching requires different interaction techniques and query analysis. As previously stated, visual searching, or content-based retrieval, operates via colors, textures, edges, shapes, and features of images or videos. One common visual search feature is to "query by example." Query by example (QBE) features allow individual keyframes to serve as a visual query. Moreover, users can select any keyframe and prompt the system to return videos with similar visual qualities. Results of a QBE search can be based on any combination of the content-based attributes identified above. Lee and Smeaton (2004) identified several other systems, including WebSEEK and SWIM, which previously implemented QBE search features.

Some retrieval systems allow users to express visual queries in the form of a sketch. Moreover, users can employ common paint tools to specify shapes, features, colors, etc. and search results, i.e. images or videos, that best match the sketch are returned. SWIM and MovEase are systems that offer a query-by-sketch search feature (Lee & Smeaton, 2004).

Histogram manipulation is rather unique visual search feature. A histogram feature presents users with the actual color map of an individual keyframe. Users can use their mouse to modify the color map and hit search; results based on the new color levels are returned. Histogram manipulation features are rare in current video IR systems. WebSEEK is one of few systems that provide such a feature (WebSEEK: Content-Based Image and Video Search and Catalog Tool for the Web, 2004).

Other querying features combine text and visual information, which is also referred to as hybrid searching. Features that employ visual and textual information have been implemented for query refinement and expansion. For example, a hybrid search feature can include 1) performing a text-based search and 2) narrowing, or expanding, search results based on visual qualities of keyframes. Hybrid search features can also be useful for tying semantic concepts to video information. A sophisticated hybrid search feature, implemented for the Físchlár system, allowed users to rate the importance of textual and visual information used in search queries (Browne et al., 2004).

3.3.2 Data Representation

In order to fully survey video IR research, issues surrounding information management and data representation must be addressed and understood. Moreover, video IR researchers should have knowledge on how interface features and functions are supported on the back end. Information used for retrieving video can be either manually created or automatically extracted. Managing both types of information brings forth a number of challenges.

3.3.2.1 Manually Created Information

There are several techniques for manually creating metadata. Metadata used by the Open Video system form Dublin Core compliant bibliographic records that also include video abstracts, descriptions, and keywords (Marchionini & Geisler, 2002). Graduate students working on the Open Video Project were actually employed to watch and manually annotate every video in the collection. The annotation performed for the Open Video Project included manually segmenting each video and assigning keywords to every shot (Marchionini & Geisler, 2002). The annotation information was organized into a database and used to implement various search and browse functions (described above).

Manually annotating video is made easier and more standardized using several existing tools, or software. Some software will import digital video and provide a customized graphical user interface (GUI) for annotating and segmenting video clips. IBM produced one video annotation software named VideoAnnEx, a publicly available toolkit (IBM Research - VideoAnnEx Annotation Tool., 2004). VideoAnnEx allows users to specify certain visual features or semantic information by employing a series of checkboxes and textboxes. For example, some of the preset features that can be selected by annotators include “man-made setting,” “fields,” “mountains,” “water,” “deer,” “human characters,” “robots,” and a wide range of others (IBM Research - VideoAnnEx Annotation Tool., 2004). VideoAnnEx allows users to annotate video content by shot, scene, or document. While manually annotating video can be a reliable and effective means for developing video retrieval systems, these efforts are often time consuming and expensive.

3.3.2.2 Automatically Extracted Information

Both text and content-based information can be automatically extracted from videos. The textual information automatically extracted from videos includes ASR and CC output, or video transcripts. ASR and CC outputs comprise all words as they are spoken throughout a video. As a result, ASR and CC outputs must be structured and organized so that they can be useful for retrieval. Organizing video transcripts has been performed using traditional IR approaches. Variations of the Okapi I and II weighting protocols have been applied to video transcripts and shown to be successful (Adams et al., 2003). Also, *tf.idf* weightings have been implemented for several video retrieval systems including one variation[47] of the Informedia (CMU) system (Hauptman et al., 2004). Simonnot and Smaï (1995) also employed *tf.idf* representation with vector space modeling.

[47] TRECVID 2003 study.

Content-based, or "low-level," information refers to the colors, textures, and shapes found throughout images and videos. Each of these features can be measured and represented using similar techniques. Analyzing colors, textures, and shapes found throughout images can be performed by counting and classifying pixels, and storing pixel sums in "bins."[48] Moreover, after each pixel is analyzed, it is classified into a bin and the sum for that particular bin is incremented by one, or the number of pixels. Analyzing images by color is typically performed using a "color-histogram," which is constructed with a three dimensional vector. The dimensions of a color histogram represent red, green, and blue (RGB) values, and each dimension contains a series of bins that signify levels of intensity for that particular color spectrum (Figure 6). One standard color histogram allocates 256 bins for each RGB dimension, representative of an 8-bit image. Other studies have evaluated 32, 128, 512, and 1024 bin color histograms. Video IR experiments using color histograms are common and have been performed by IBM, Carnegie Mellon, Dublin City University, and a number of other researchers.

Figure 6: Example of color histogram (3 * 2 Array/Vector).

	Bin.1	***Bin.2***	***Bin.3***	***Bin.4***	***Bin.5***	***Bin.6***	***Bin.X***	***Bin.Y***	***Bin.Z***
Red	564	6512	327	89465	651	951	6542	78561	561
Green	86451	5212	8754	5641	865	54123	213	321	886
Blue	684	321	684	864	156	564	357	5251	321

y axis

x axis

Measuring texture and edge qualities can be performed similarly to using a color histogram, i.e. counting pixels and incrementing bin values. However, the data structures

[48] Each bin is associated with a descriptive, or visual, trait.

of color, texture and edge histograms typically vary between three dimensional vectors (for a color histogram) and one dimensional vectors. The bins associated with texture and edge histograms represent gradients, directions, and/or image “flow” characteristics (Amir et al., 2003).

Although these are standard techniques for extracting and representing content-based information, histogram outputs must be further analyzed and processed in order to make useful for retrieval. The research groups discussed throughout this section have employed predominately unique approaches for analyzing histogram statistics (the reader is recommended to refer to the original source for further details). In addition, researchers from this study previously[49] implemented color histogram extraction applications using Java’s Advanced Imaging (JAI) API where color levels for various video shots[50], i.e. keyframes, were measured. Color levels of each keyframe were then compared with all other keyframes and a color similarity value for each unique combination was computed using the sum of absolute bin differences. Nevertheless, the creation and evaluation of original image similarity computations is beyond the scope of this study and will not be explored thoroughly.

3.4 Chapter Summary

This chapter has demonstrated that visual and non-visual information seeking studies are pertinent for analyzing video searching behaviors. Discussion of searching behaviors

[49] Image analysis was performed for the TRECVID 2004 study.

[50] A (RGB) 32-bin color histogram was used to analyze color levels in the TRECVID 2004 keyframes.

addressed visual and textual search tasks from specific domains and professional communities. Video retrieval systems deriving from renowned research groups, including IBM, Carnegie Mellon, Dublin City University, and the University of North Carolina, were surveyed throughout this chapter. The systems developed by each of these groups helped signify important advances in video IR research including systems functionality, query modeling, interface design, information management, and evaluation.

Chapter 4

Methodology

To investigate the problems raised in this study, a series of search experiments has been carried out. Each experiment was designed to reflect actual video searching scenarios and tasks facing science educators. Several system variants were developed to measure associations between the use of various interface features and functions, domain-specific search tasks, and systems effectiveness. The experimental design, in general, incorporated both quantitative and qualitative methods for collecting and analyzing data. Results from the actual experiments depict user interaction with video search systems, task performances, and other user judgments relating to interfaces and search tasks. The terms "user" and "subject" both refer to the participants of this study and are used interchangeably throughout this chapter.

4.1 Experimental Design

This study has incorporated a task-centric approach to evaluating interface features and functions for video retrieval systems. Factors associated with video retrieval systems and search tasks have been analyzed according to influences on science educators' searching behaviors. A variety of interface features and functions were implemented as part of several search systems, and the effect of each variant on task performance was evaluated. These searching experiments required each subject to complete a certain number of search tasks using three different system variants. Results from the search experiments were analyzed using objective and subjective evaluation. Objective results were used to analyze interaction behaviors, task performances, and subjects' demographics. Subjective evaluation measured other factors, such as the subjects' assessment of system effectiveness, interface functionality, task representation, and prior knowledge.

4.1.1 Goals and Parameters

The overall goal for this study is to determine how video retrieval systems, including interface features and functions, can be designed to support domain-specific search tasks. First, the authors believe that frameworks, or standards, for developing video retrieval interfaces are needed and – in order to begin establishing such standards – researchers may first want to explore video IR problems from the perspective of search tasks. Challenges associated with this approach bring forth a number of different experimental parameters and factors. The different parameters and factors developed for this study are presented in Table 1.

Table 1: Experimental details.

Parameters	**Data Collection**		**Data Analysis**	
	Factors measured	Instruments or techniques	Related parameters	Measurement and comparison methods
User Demographics and Prior Knowledge	Experience with search tools, science education research, and video searching, familiarity with tasks	Pre-test and Post-Search questionnaires	Task representation, performance, user interaction, system functionality	Descriptive statistics, mean comparison and correlation tests
User Interaction with Interface Features	Use of keyword search, video browse, results browse, promote search, video details, visual search, steps, time	Observation, during-test statistics, server logs	Task representation and familiarity	Descriptive statistics, mean comparison and correlation tests
System Functionality and Performance	Interface support for task completion, usability of system, "learnability" of system, task completion, errors	Post-search and post-experiment questionnaires, interviews, observation, during-test statistics	System variant, performance, task representation, users' prior knowledge	Descriptive statistics, mean comparison and correlation tests, performance coding
Task Representation	Accuracy of "real" tasks, task classification	Post-search questionnaires	User Interaction, task performance, system functionality, and prior knowledge	Descriptive statistics, task coding, mean comparison and correlation tests

One such set of parameters of this study pertained to user details and prior knowledge. Considering that this study employed a domain- and task-centric approach to evaluating video retrieval systems, it was imperative that the subject pool consist of users who have experience in Science Education. Important factors corresponding to this subject group included: experience with general (electronic) search tools, performing science educational research, searching video retrieval systems, or video digital libraries, and how familiar the subjects' were with the content of each task.

Other experimental parameters created for this study involved searching behaviors, system effectiveness, and performance. The researchers of this study evaluated interaction patterns, including the number of steps and errors[51], task times, and use of individual interface features. In addition, systems- and performance-related factors gauged interface effectiveness, system usability and "learnability"[52], and task completion.

Task-oriented parameters were also evaluated as part of this study. These parameters primarily assessed task representation and classification. Analyzing task representation supported the validation of the experimental tasks. Moreover, it was important that the researchers assess how reflective the experimental tasks were of "real" tasks facing science educators. The search tasks developed for this study were also categorized so search experiments could evaluate systems performance and user interaction across different task types.

4.1.2 Experimental Systems

The search experiments evaluated several different variants of the ViewFinder video retrieval system. The purpose behind ViewFinder is to provide a general interface that

---

[51] Errors, as performed by the subjects, were defined for this study as unmistakable incorrect queries or misuse of any interface feature or function.

[52] "Learnability" refers to the user's capability of learning to use the system.

can be applied to quickly implement and evaluate different interface features and functions. Results from ViewFinder have been previously evaluated throughout several years of the TRECVID search task. The versions of ViewFinder implemented for this study were designed to search and browse the NASA K – 16 Science Education Programs. Chapter 5 presents a complete description of the ViewFinder experimental systems, including system components and interface features and functions.

Throughout this chapter, system details will be discussed on a general level. The interface features and functions evaluated as part of this study reside on three basic levels[53]: keyword searching, video browsing, and querying by example[54]. A number of interface features and functions correspond to each of these general components.

Table 2: System variants.

Experimental Systems	Features and Functions
Variant 1	Keyword Search, Query By Example
Variant 2	Video Browse, Query By Example
Variant 3	Keyword Search, Video Browse, Query By Example

A total of three different variants was implemented for these experiments (see Table 2). Each system variant incorporated a different set of interface features and

[53] When evaluating video search functions, factors surrounding search results display must also be addressed.

[54] Query by example (QBE) features were based on hybrid retrieval, i.e. both textual and visual information were used to form queries and search video data.

functions. System variant 1 evaluated keyword searching[55] and querying by example, while variant 2 tested video browsing together with querying by example, and variant 3 incorporated all three functions, i.e. the "full" system variant. (See Chapter 5 for complete description of these experimental systems, including the development of all features and functions evaluated by this study.) By excluding certain features throughout different system variants, researchers were able to isolate and test the effect of various system parameters across task characteristics and user behaviors, and validate any observed associations or differences.

4.1.3 Experimental Details

The researchers of this study considered many different issues when developing the experimental design. Some of the issues that had to be addressed throughout experimental planning included: assembling the subject sample and test settings, and designing the experimental search tasks and search runs[56].

Subject Recruitment

For this particular study, it was imperative that the subjects be recruited directly from Science Education, including teachers and education students. The experiments developed for this study strive to reflect actual needs and tasks facing science educators.

---

[55] The keyword search retrieves video based on a variety of textual information, ranging from video transcripts to manually assigned descriptions and keywords.

[56] A "search run" refers to one full search experiment where sequencing and ordering of tasks, system variants, and subjects have been resolved.

This study recruited current and former science teachers from K – 12 schools, and Science Education majors from the School of Education at Indiana University, Bloomington. To recruit these students, the researchers contacted various school administrators and Science Education instructors[57] and sought permission to sign-up interested participants. The researchers of this study recruited a total of 28 subjects. Each subject was compensated $10 per hour of participation. Researchers did not selectively choose or filter subjects; subjects were selected on a first-come basis.

Task Creation

A series of search tasks was developed for the experiments. The search tasks were designed to reflect "real" tasks facing science educators, such as finding video materials to support lab activities or classroom lectures. Several tasks were previously created and evaluated as part of a pilot study. Experimental tools and results from the pilot study are presented in Appendices D through M. Subjects recruited for the pilot study responded positively to the representation, or accuracy, of the search tasks, and therefore validated our task creation process. The actual search tasks developed and evaluated by the pilot study are found in Appendix D.

The search tasks of the formal study (Appendix M) were developed to search and browse the NASA K – 16 Science Education Programs, and encompassed a wide-range of educational topics. A total of eight search tasks was developed for this study. All of the search tasks have been classified into certain categories. The search tasks were

[57] The e-mail asking for instructors' permission to recruit students is located in Appendix B.

categorized based on the number of sub-tasks required for task completion and the use of visual and textual clues[58]. Definitions of the task types developed for this study include:

- *Easy task* – a task consisting of one sub-task, using either visual or text-based information. There can be text-based easy tasks and visual easy tasks.
- *Complex task* – a task consisting of multiple sub-tasks and based on either textual or visual information. There can be text-based complex tasks and visual complex tasks.
- *Combination task* – a task consisting of one sub-task, which involves (both) textual and visual information.
- *Combo-complex task* – a task consisting of more than one sub-task, and involves both textual and visual information.

The eight tasks developed for this study consisted of two easy[59], complex[60], combination, and combo-complex tasks, apiece.

Experimental Structure and Environment

The distribution of search tasks across different system variants and individual subjects is presented in Table 3. Table 3 exhibits that each subject completed a total of six (out of eight) search tasks. The search tasks were ordered[61] in a way to compensate for a learning curve; however, the tasks that involve the use of System variant 3 – the full

---

[58] The task categorizations employed by this study were rationalized by the researchers' need to explore multiple variations and characteristics of search tasks, and to examine their influence(s) on other experimental parameters.

[59] The two easy tasks included one text-based easy task and one visual task.

[60] The two complex tasks included one text-based complex task and one visual tasks.

[61] See Table 4 for the exact order, or sequence, of the search tasks.

system – were given at the end of the search run, after users were exposed to systems that excluded certain features (see Table 4). The experimental design also indicates that each task type was performed a total of 42 times, with equal distribution across system variants and textual and visual types of search tasks. The equal distribution of experimental parameters enabled researchers to compare across different factors, including system variants, subjects, and task types.

Table 3: Experimental design; Distribution of tasks and system variants among subjects.

Legend: t = task, s = subject, v = system variant, et = easy task (textual), ev = easy task (visual), cxt = complex task (textual), cxv = complex task (visual), cm = combination task, cc = combo-complex task.

		et	ev	cxt	cxv	cm	cm	cc	cc
		T1	T2	T3	T4	T5	T6	T7	T8
g1	s1	v1	v2	v3		v1	v2	v3	
g1	s2	v1	v2	v3		v1	v2	v3	
g1	s3	v1	v2	v3		v1	v2	v3	
g1	s4	v1	v2	v3		v1	v2	v3	
g1	S5	v1	v2	v3		v1	v2	v3	
g1	S6	v1	v2	v3		v1	v2	v3	
g1	s7	v1	v2	v3		v1	v2	v3	
g2	s8	v2	v3		v1	v2	v3		v1
g2	s9	v2	v3		v1	v2	v3		v1
g2	s10	v2	v3		v1	v2	v3		v1
g2	s11	v2	v3		v1	v2	v3		v1
g2	s12	v2	v3		v1	v2	v3		v1
g2	s13	v2	v3		v1	v2	v3		v1
g2	s14	v2	v3		v1	v2	v3		v1
g3	s15	v3		v1	v2	v3		v1	v2
g3	s16	v3		v1	v2	v3		v1	v2
g3	s17	v3		v1	v2	v3		v1	v2
g3	s18	v3		v1	v2	v3		v1	v2
g3	s19	v3		v1	v2	v3		v1	v2
g3	s20	v3		v1	v2	v3		v1	v2
g3	s21	v3		v1	v2	v3		v1	v2
g4	s22		v1	v2	v3		v1	v2	v3
g4	s23		v1	v2	v3		v1	v2	v3
g4	s24		v1	v2	v3		v1	v2	v3
g4	s25		v1	v2	v3		v1	v2	v3
g4	s26		v1	v2	v3		v1	v2	v3
g4	s27		v1	v2	v3		v1	v2	v3
g4	s28		v1	v2	v3		v1	v2	v3

Table 4: Experimental design; Task ordering.

	et	ev	cxt	cxv	cm	cm	cc	cc
	T1	T2	T3	T4	T5	T6	T7	T8
s1	1	2	5		3	4	6	
s2	2	3	6		4	1	5	
s3	3	4	5		1	2	6	
s4	4	1	6		2	3	5	
S5	1	2	5		3	4	6	
S6	2	3	6		4	1	5	
s7	3	4	5		1	2	6	
s8	4	5		1	2	6		3
s9	1	6		2	3	5		4
s10	2	5		3	4	6		1
s11	3	6		4	1	5		2
s12	4	5		1	2	6		3
s13	1	6		2	3	5		4
s14	2	5		3	4	6		1
s15	5		3	4	6		1	2
s16	6		4	1	5		2	3
s17	5		1	2	6		3	4
s18	6		2	3	5		4	1
s19	5		3	4	6		1	2
s20	6		4	1	5		2	3
s21	5		1	2	6		3	4
s22		2	3	5		4	1	6
s23		3	4	6		1	2	5
s24		4	1	5		2	3	6
s25		1	2	6		3	4	5
s26		2	3	5		4	1	6
s27		3	4	6		1	2	5
s28		4	1	5		2	3	6

A variety of experimental settings were organized by the researchers. Each search experiment was held in a reserved room – either onsite or at the School of Library and Information Science at Indiana University, Bloomington – where only the experimenter and subject were present. Holding the experiments in a reserved room helped eliminate distractions that accompany public computer clusters. The subjects performed each search task on a standard laptop PC where the ViewFinder systems were accessed through the Web using a Java enabled browser. Prior to the beginning of each search experiment, users read and signed an informed consent statement, see Appendix L, and were given a five-minute tutorial on the system. A five-minute time limit was set for each search experiment, but the subjects were encouraged to end each task whenever they deemed it complete, or success as unattainable. (Methods for collecting experimental data, including task completion times, will be presented in section 4.14 *Data Collection*.)

4.1.4 Data Collection

A variety of data collection tools, encompassing both quantitative and qualitative methods, were created for this study. Questionnaires[62] [63] were used to gather data about subjects, tasks, interface features and functions, and retrieval systems. "During-test" data, which was collected by observing subjects perform each search experiment, measured certain factors related to performance and interaction, including task completion ratios, completion times, steps and errors. After each search experiment, the researchers posed several interview questions and allowed each subject to elaborate on their experience. Server logs were also collected, which enabled researchers to revisit the search experiments and validate the interaction data tallied throughout the during-test observations. Screengrabs were taken as part of the pilot study; however, server logs were shown to be sufficient means for validating during-test data for the formal study and screengrabs weren't utilized.

---

[62] The questionnaires developed for the pilot study are presented in Appendices G through G.

[63] All questionnaires originated from the TREC Interactive Search Task and were modified for this particular study (TREC-9 interactive searching study, 2004).

A series of questionnaires, three in total, was administered at different intervals of the search experiments. The first questionnaire, the pre-experiment questionnaire, was given to the subjects prior to any search task or systems tutorial. The pre-experiment questionnaire, found in Appendix N, was important because it helped assess user-related characteristics and factors. General information about the subjects, including education level, age, current occupation, and academic major, were all collected using the pre-experiment questionnaire. Next, several questions measured the technological skills of each subject. The subjects were asked to rate their experience level with using point and click interfaces, online commercial databases, Web search engines, and video retrieval systems. Subjects were also asked to rate their experience with other tasks related specifically to Science Education, including developing lesson plans, assignments, lab activities, and lecture materials, and incorporating video into classroom activities.

A post-search questionnaire, found in Appendix O, was another tool for collecting experimental data. Post-search questionnaires were administered after each search task; therefore, subjects completed a total of six questionnaires. The experimental parameters measured by the post-search questionnaire involved task representation, system functionality, and performance. Subjects were asked to rate their familiarity with the content of each task, how representative the task was of a “real” search tasks facing science educators, and how easy it was to search each task. Next, subjects reported their satisfaction with system functionality, interface features and functions, and search results. Subjects were then asked to self-assess the extent to which they successfully completed each search task.

A third, and final, questionnaire was administered after the completion of all search tasks. The post-experiment questionnaire, presented in Appendix P, primarily focused on general usability and "learnability" of the ViewFinder system. Subjects were asked to rate their overall satisfaction with system functionality and interface features and functions. In addition, the post-experiment questionnaire allowed subjects to estimate the ease of assessing video content and the necessity for playing the video (see Appendix P for exact questions included on this questionnaire).

Researchers also monitored the search experiments and manually recorded all actions performed by the subjects. Moreover, the number of times each interface feature was selected and used by a subject was recorded. These results, referred to as during-test results, allowed researchers to calculate the total number of steps and errors performed for each task. Starting and stopping times, or the overall completion times, were also recorded. The experimental datasheet used to record during-test results is presented in Appendix K.

A series of interview questions was also posed to each subject after the completion all search tasks. Generally speaking, subjects were asked to elaborate on which interface features and functions were found most useful and which features were most confusing, or ineffective. All interview sessions were recorded on a PC using audio recording software and later transcribed. The interview questions used for this study are presented in Appendix Q.

As previously mentioned, server logs were collected for each search task. Moreover, the ViewFinder system was programmed to output all actions performed by the subjects to a text file. The server logs enabled researchers to revisit the search experiments to validate the during-test results.

4.1.5 Data Analysis

This study comprised a variety of data analysis techniques. First, objective analysis assessed during-test results, task performances, and user demographics. Subjective analysis was also performed. Subjective results depict subjects' assessment of all interface features and functions, and other systems- and task-centric factors. Mean comparisons and correlation analysis were performed across many experimental factors. It is important to note that methods for analyzing individual parameters are presented first and discussion of correlation and mean comparison are presented toward the end of this section.

4.1.5.1 Objective Analysis

Objective results were first used to analyze user demographics, in a general sense and in relation to technological skills, previous experiences, and training. Objective analysis also measured task completion rates, (task) durations, and the number of steps and errors performed throughout each search experiment.

Subject Demographics

The demographics of interest to this study primarily focused on professional and academic backgrounds, technological skills, and the use and application[64] of digital video. A pre-experiment questionnaire (located in Appendix N) was used to gather demographic information. On a 5-point scale[65], subjects were asked to rate their skills and experiences. This assessment enabled researchers to perform various statistical analyses to characterize subjects and groups. Means, minimums, maximums, range(s) and other descriptive statistics were all computed. Individual subjects were categorized and sub-samples, consisting of those with similar skills and experiences, were formed and assessed further. Factors related to more general attributes, including sex and age of the subjects, did not comprise any additional variables. Further discussion of this analysis is presented at a later section.

During-Test Results

Data collected through the monitoring of each search task was analyzed to produce “during-test” results. As previously mentioned, during-test data primarily focused on task completion ratios, task durations, user actions, and steps and errors. Analyzing during-test results helped researchers identify tasks that were completed more successfully and those completed less frequently. Researchers were also able to recognize which tasks subjects were willing to invest more time and energy completing. Descriptive statistics were sufficient for drawing certain conclusions about task difficulty.

---

[64] Application of video refers to how teachers incorporated or applied video into classroom settings for instructional purposes.

[65] On each of the questionnaires, “5” indicates highly agreeable, and “1” indicates unfavorable.

The researchers were able to compare and rank individual tasks according to these measures. Throughout this analysis, researchers compiled task performance data with considerations[66] for task outcome, i.e. subjects' self-assessment of task completion as collected using the post-search questionnaire (see Appendix O).

During-test results also reflected usage patterns of specific interface features and functions. For example, researchers were able to distinguish between tasks that required higher levels of keyword searching and tasks where users preferred to browse categorically. Computing composite usage statistics also helped researchers estimate preferences for interface features and functions in general, or across all search experiments. These results were also analyzed using various frequencies and descriptive statistics.

The during-test analysis, described throughout this section, was performed on a couple of different levels. Data analysis entailed task completion and performance assessment using descriptive statistics and ranking, and measuring user interactions and usage patterns across specific interface features and functions and system variants. Each of these techniques was applied in order to analyze results across individual tasks and specific task types.[67] Moreover, during-test results were analyzed in line with different task classifications, which indicated differences in task complexity, user interactions, and interface preferences.

[66] Such consideration helped control for tasks that users gave up on.

[67] Task types include complex, combination, and combo-complex tasks; each contains various levels of textual and visual needs and clues.

4.1.5.2 Subjective Analysis

Subjective factors were created to assess subjects' judgments on system functionality, interface usefulness, and task representation. Using several questionnaires[68], subjects rated their satisfaction with system effectiveness, individual features and functions, and the usability and "learnability" of the ViewFinder system. Subjective results also measured users' perceptions about each search task. The questionnaires used to gather subjective data were based on a 5-point scale, where "5" indicated extreme agreeability and "1" symbolized disagreeability. Similar to the statistical methods previously described, when analyzing each factor independently, a series of descriptive statistics and rankings were useful for evaluating system functionality, interface support, and task characteristics. These statistical results enabled researchers to identify supportive interface features and functions and realistic search tasks.

Several interview questions, posed after each experiment, were used to follow-up on questionnaire responses and during-test results. Users were asked to elaborate on which interface features and functions they found to be most useful and which ones they found most confusing. Analysis of the interview responses included frequencies and rankings for the individual interface features and functions specified as supportive and ineffective.

[68] Subjective data was gathered in the post-search and post-experiment questionnaires.

4.1.5.3 Examined Relationships

As presented in Table 1, many relationships among individual factors have been analyzed. Comparisons between individual factors were performed using tests of correlation and mean comparisons. More specifically, the statistical analyses that were performed for this study included independent and dependent samples T-Tests and correlation analysis. Statistical significance for each of these tests was the 0.05 and 0.01 standard levels.

To highlight a few examples of the tests performed for this study, let's first consider comparisons between different subjective parameters. One relationship that was explored included correlating subjects' familiarity with search tasks, i.e. prior knowledge, with their perceptions of interface effectiveness. Collectively, these factors helped researchers analyze how domain expertise may influence user's assumptions about interfaces and retrieval systems. Relationships between interface designs and task familiarity were evaluated across individual tasks and task types. From examining the subjective factors, as presented in Table 1, it is apparent that many comparisons are possible.

Associations between objective parameters were also explored. For example, it was important to test for correlations between task performances and interface usage. Many other relationships between objective parameters were also explored (see Table 1). More specifically, two examples of these relationships evaluated from this study included (a) how task completion ratios were correlated with the use of a keyword search, and (b) how task durations differed across task types. Similar to the discussion above, correlation tests and mean comparisons between objective results were performed as they relate to individual tasks and task classifications.

Finally, this study examined relationships among factors that spanned objective and subjective analysis. Researchers were interested in discovering if task performance was related to task characteristics and subjects' prior knowledge, i.e. task familiarity. The methods used for this study also allowed researchers to specifically examine how subjects performed on more general search tasks and tasks that were deemed more representative of Science Education. Another comparison between subjective and objective results included correlating subjects' perceptions about system functionality with task completion ratios. Again, readers should remember that many more relationships were explored throughout this analysis. Experimental results are presented in Chapter 6.

4.1.6 Discussion

What do these methods and results mean for future video retrieval studies? The data analysis techniques performed throughout this study were important for a number of different reasons. First, results helped researchers understand relationships between the interface features and functions of video retrieval systems and domain-centric, i.e. science educational, search tasks. Moreover, this study recognized certain interface features and functions that were supportive for science educational video retrieval systems. Results from this study may help researcher implement and eliminate features for similar (educational) systems. These experiments may also help facilitate the development of standards and principles for video retrieval research.

Other findings from this study provide further understanding of the search tasks for video retrieval research. By developing the experimental methods as described in this chapter, researchers first classified video search tasks using different attributes, including the number of sub-tasks and the inclusion (or omission) of visual and textual information. A domain-centric approach to evaluating user interfaces may also help researchers better understand search tasks, as task analysis is inherent in studies that isolate and explore domain. Future task analysis may help produce other task categories, which may not have been previously considered important for Science Education. The results from this study assert how understanding video search tasks is important for similar studies. Currently, video retrieval research lacks foundations for exploring the influences of task characteristics on user interaction, searching performance, and retrieval.

This study may also help video retrieval researchers analyze users from a particular domain. One important dimension of this study assessed the differences among sub-groups within the sample. Moreover, future studies can enable researchers to begin understanding how certain skills, domain expertise, and perceptions may influence interaction with video retrieval systems, and how findings can be applied for designing interface features and functions.

4.2 Methodology Summary

This chapter has presented a methodology for exploring domain-specific problems in video retrieval research. This discussion has detailed the experimental parameters and factors that have been evaluated. A total of three system variants were designed to investigate the effectiveness of different interface features and functions. Data collection tools, including questionnaires, observation methods, and interviews, were employed throughout many search experiments. Data analysis also helped determine the implications for system designs and experimental methods. The experimental factors and parameters encompassed task performance, domain expertise, system functionality and usability, and user interactions.

Chapter 5

Experimental System

An experimental system was developed to evaluate the influences and associations of Science Education video search tasks on user interaction so that future efforts may be able to design more effective video retrieval systems and user interfaces. To implement such a system, the researchers assembled a video dataset, which was produced for one particular audience, i.e. science teachers. The dataset used for this study was gathered with the support of the Open Video Project[69] (http://www.open-video.org) at The University of North Carolina (UNC) – Chapel Hill. UNC researchers granted access to the NASA K – 16 Science Education Programs and other metadata generated for the Open Video search system. A web-enabled user interface, database, and client-server communication programs were all developed for this study.

ViewFinder, a video retrieval system, was implemented for this study. This particular version of ViewFinder was used to measure the effectiveness and use of

[69] The Open Video search system was reviewed in sections 2.5 and 3.3.

interface features and functions when supporting domain-specific (science educational) search tasks. ViewFinder is a Java-driven system supported by a backend Oracle database. Results of ViewFinder were previously evaluated through several years of the Text REtrieval Conference's Video Retrieval Evaluation, or TRECVID.

5.1 Dataset

The NASA K – 16 Science Education Programs were produced by NASA with governmental funds, and are open to the general public. The collection contains newer programs, production dates ranging from 2000 to 2006, and high quality video. The video collection comprises several NASA series including *NASA Connect*, *NASA SciFiles*, *NASA Why?Files*, and *Destination Tomorrow*. These videos are in full color and English. Overall, this video collection spans many subjects ranging from weather forecasting to archeological explorations. A few sample programs from the *NASA Connect* series includes:

- *Ancient Observatories*
- *Better Health from Space to Earth*
- *Good Stress*
- *The Measurement of All Things*
- *Tools of the Aeronautic Trade*

The "physical" characteristics of the dataset are of such: there are currently 54 full video programs that total approximately 31 hours of content. The size of the collection, depending on video format, ranges from 71 GB for MPEG-2 to 6 GB for

QuickTime[70]. The 54 video files within this collection consist of full programs, including 30 and 60 minute documentaries.

5.2 Metadata and System Architecture

An assortment of metadata corresponding to the NASA K – 16 Science Education Programs was manually created by Open Video researchers. Researchers at Open Video developed and assigned Dublin Core bibliographic records for all full programs. The structure of an Open Video record is presented in *Figure 7: Open Video Record*. UNC researchers willingly shared this manually generated metadata, for all video programs, with this study.

Several of these manually annotated fields were used to develop different features of the ViewFinder system. For example, the researchers of this study implemented a version of ViewFinder to browse the video collection by title, as extracted from the Open Video records, which then enabled the evaluation certain interface features based on manual indexing. The interface features and functions of the Open Video system were surveyed in sections 2.5 and 3.3.

In order to evaluate other interface features, separate from the current Open Video system, researchers collected, extracted, organized, and indexed additional metadata. On a textual level, transcripts of the NASA K – 16 Science Education Programs were indexed and organized for retrieval. Retrieving video through transcript was shown

[70] The exact size for all video formats, across the entire collection, include 71GB MPEG-2, 42GB MPEG-1, 14 GB REAL, 12GB MPEG-4, and 6GB for QuickTime.

effective through past years of TRECVID evaluation. Previous versions of ViewFinder incorporated a transcript-based search feature where search results were generated and assessed for TRECVID participation. Transcripts for the NASA video collection were generated by NASA and shared with researchers from UNC (and subsequently this study). As reviewed in Chapter 3, designing features that search video by transcripts pose many challenges for systems developers. The current Open Video web system does not include a transcript-based function for searching or browsing the NASA K – 16 Science Education Programs.

Figure 7: Open Video Record.

VIDEO INFORMATION
TITLE
DESCRIPTION
YEAR
GENRE
KEYWORD
DURATION
COLOR
SOUND
AMOUNT OF MOTION
LANGUAGE
SPONSOR
CONTRIBUTING ORGANIZATIONS
TRANSCRIPT AVAILABLE
COPYRIGHT STATEMENT
DESCRIPTION

Keywords deriving from the transcripts, i.e. the actual spoken words throughout the videos, were organized and accessible for retrieval. Organizing transcript data was accomplished by weighting each keyword. *Tf•idf* weighting, modified[71] to reflect a video's structure, provided effective means for indexing the video transcripts. Weighting transcript terms, along with video segmentation, allowed subjects to retrieve results on a "finer[72]" level, rather than by whole document, or video, alone.

Other additions to ViewFinder included various content-based, or visual, searching features. As reviewed in Chapter 3, there are many challenges associated with content-based retrieval. Content-based analysis strives to produce meaningful metadata for video documents by automatically processing visual properties. Many content-based retrieval efforts focus on drawing semantic information from the visual elements of images or videos.

The content-based retrieval features, implemented for this study, included searching by color, shape, texture, and the combination of each. Developing features to search by these visual attributes required a great deal of image processing and information management. Techniques used to develop these visual search features differed from previous discussions of content-based retrieval, where histograms were highlighted as means for extracting and processing visual qualities of images or videos. (See Chapter 3 for review visual histograms, including RGB color histograms.) For this particular study, visual analysis, i.e. the extraction of colors, shapes, and textures, was performed using Oracle's interMedia, a pre-existing tool (interMedia, 2007). Oracle interMedia is a database tool developed to manage and process multimedia information,

[71] IDF was computed using the number of minutes per video and the number of minutes where the terms appeared. TF represented the number of times the word appears in a minute.

[72] Shorter, or smaller, pieces of video. For this particular study the videos and transcripts were segmented by minute.

particularly images and videos. In addition, the Oracle interMedia package also includes a Java API, which implements various classes that can analyze the similarity of different images according to visual attributes (interMedia, 2007). For this study, researchers imported images into an interMedia table and created applications – using the interMedia Java API – that produced similarity scores based on Oracle's computations[73]. When analyzing images with interMedia, similarity classes are designed to output statistics that range from 0.0 to 100. A score of 0.0 indicates high similarity among different images, and a score of 100 signifies a vast difference.

As previously mentioned, the interMedia Java API can generate similarity statistics according to any combination of the color, shape, and texture qualities of an image. When calling methods that evaluate similarity scores, researchers have the ability to weight the different visual attributes as parameters. Researchers are able to weight each visual attribute between a score of 0.0 (low) and 1.0 (high). Any combination of these weightings can be applied across any or all attributes for assessing visual similarity. For this particular study, similarity statistics were generated using high (1.0) scores for each individual visual attribute, i.e. color, shape, and texture, and all of them together, i.e. all visuals.

In the end, researchers were given similarity statistics for each unique comparison of all keyframes in the collection, which were then stored in the Oracle database. Each similarity measure could be identified by the unique minute IDs, a composite key. When a subject performed a visual search, all matching keyframes, or minutes, with similar

[73] The algorithms, or formulas, implemented by Oracle for measuring visual similarity across images are not thoroughly documented.

visual attributes were returned to the client and ranked according to Oracle's similarity score.

This approach for content-based retrieval was applied to the NASA video collection and different visual search functions, based on color, shape, and texture values, were incorporated into ViewFinder. In addition to computing visual similarity scores, other challenges were associated with task. One challenge was that the researchers were required to extract certain keyframes to represent the minute-long segments of the NASA video. For this particular set of the NASA Programs, Open Video researchers had previously generated close to 84,000 keyframes[74] which were selected using a frame incrementing technique[75]. Such a technique can be useful; however, it doesn't necessarily support generating keyframes to represent specific minutes of video. For example, selecting keyframes at frame number *X*, *X+100*, *X+200* ... is not useful when extracting keyframes to represent of *minuteX*, *minuteY,* and *minuteZ,* because minute boundaries[76] are not measured at exactly 100 frames and too many keyframes accumulate. Many other challenges accompanying image analysis, i.e. processing speed, power, and mass storage capabilities, were all explored throughout previous ViewFinder studies.

---

[74] The keyframes extracted from the NASA dataset were all shared with the researchers of this study.

[75] The selection of keyframes at various frame increments, i.e., choosing keyframes at frame number X + 100.

[76] Video processing techniques, such as shot boundary detection and keyframe extraction, are beyond the scope of this study; however, viable solutions to these challenges were resolved for the experimental versions of ViewFinder.

ViewFinder is a web-accessible search system. Communication between the client and backend Oracle database is maintained through the use of Java Servlets. The Servlets implemented for ViewFinder are responsible for receiving and processing client-side queries, formatting SQL queries, transmitting to the database, retrieving and processing results from the database, and transmitting results back to the client. A model of ViewFinder's basic architecture is presented in *Figure 8.*

Figure 8: Architecture of ViewFinder prototype system.

5. 3 Interface Features and Functions

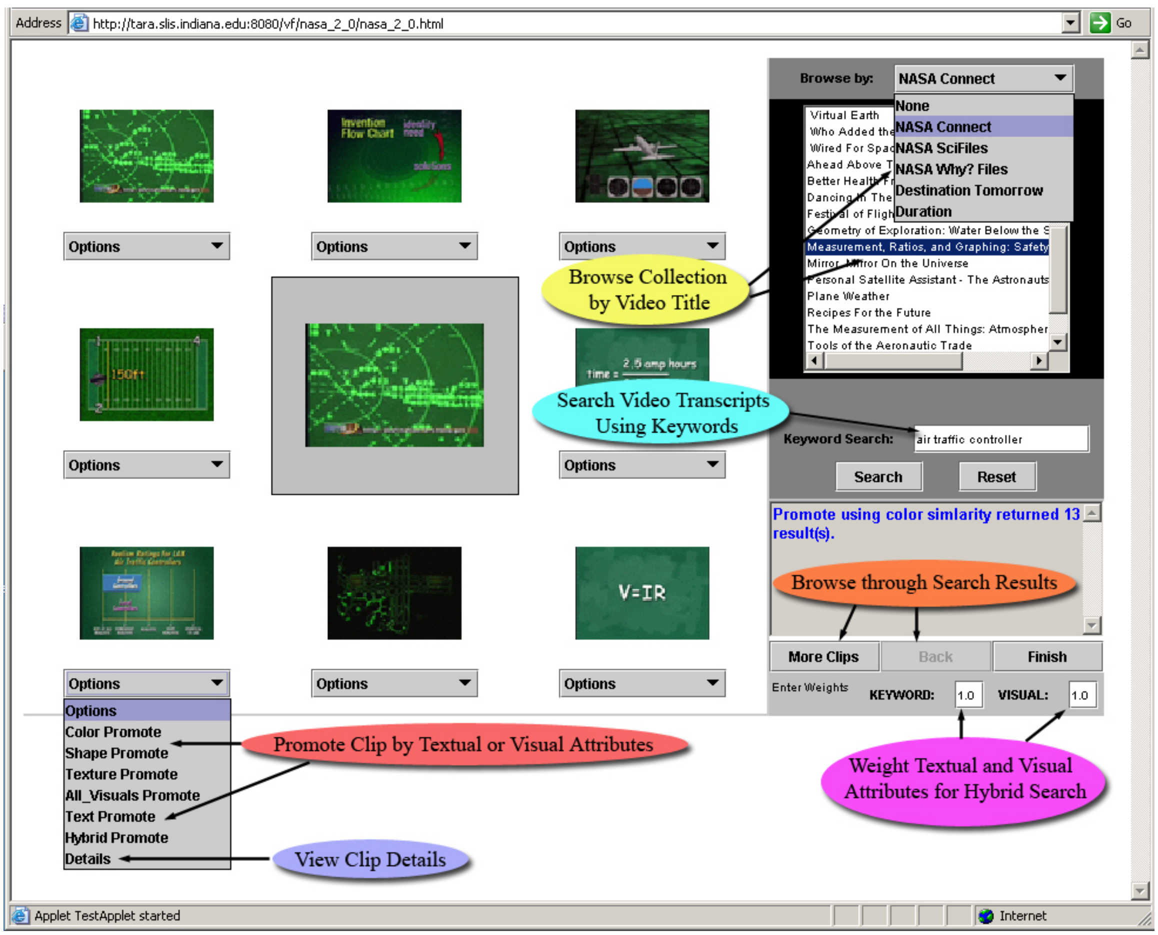
It is important to remember that there are three different versions of ViewFinder described throughout this paper. ViewFinder has been implemented for multiple years of TRECVID evaluation, a pilot study preceding these experiments, and the formal study. Each version of ViewFinder contained some overlapping features and functions. Table 5 exhibits the different interface features and functions implemented for three versions of ViewFinder, including the TRECVID 2004 system and the NASA pilot and experimental systems. The system discussed in this section is the experimental version of ViewFinder implemented for the formal study.

Table 5: Interface features and functions from different versions of ViewFinder.

	TRECVID 2004 System	NASA Pilot System	NASA Experimental System
Keyword Search of manual annotations, i.e. text		●	
Transcript Search	●		●
'OR' search by default	●	●	●
'AND' search function		●	●
'NOT' search function			
Phrasal searching			
Promote Search using keywords (QBE)	●	●	●
Promote Search using color similarity (QBE)	●		●
Promote Search using texture similarity (QBE)			●
Promote Search using shape similarity (QBE)			●
Promote Search using All Visuals similarity (QBE)			●
Hybrid Promote Search (QBE)	●		●
Adjustable weighting of Promote Search attributes	●		●
Browse by Date	●		
Browse by Source	●	●	●
Browse by Duration		●	●
Browse by Title		●	●
Details feature	●	●	●
Use of Keyframes	●	●	●
Search results ordered by video (id number)		●	
Ranked search results (keyword weighting)	●		●
Ranked search results (visual qualities)	●		●
Ranked search results (hybrid)	●		●
Retrieval by shot	●		
Retrieval by minute			●
Retrieval by segment (manually segmented)		●	
Retrieval by video	●	●	●

The graphical interface of the ViewFinder experimental system (presented in Image 3) is a web-accessible Java Applet. The interface features and functions were developed using Java Swing. The ViewFinder interface[77] is made up of two primary panels including the results panel, a grid layout, and the search panel, a border layout. The results panel, located on the left hand side, contains several features. First, it displays keyframes for the search results. The results panel can display up to eight search results on a page and ranks search results by relevance, i.e. top-left to bottom-right is the descending order.

Image 3: Screenshot of ViewFinder developed for NASA collection.



[77] The ViewFinder interface, as a whole, was implemented as a border layout.

The results panel also contains the Promote and Details functions. Different Promote functions and Details can be selected from the drop down menus located below each of the eight search results; each menu corresponds to the keyframe placed directly above it. Promote executes a new search. A Promote search, or to query by example (QBE), performs similarly to a “More like This” search feature offered by several popular search engines. That is, information associated with the selected, or promoted, result is used to formulate a query. The experimental version of ViewFinder, designed to search and browse the NASA K – 16 Science Education Programs, incorporated a color, shape, texture, all visuals, textual, and hybrid Promote function. Therefore, users could QBE using any of these visual or textual attributes. When performing a hybrid Promote search, users were also allowed to weight keyword and visual values, and add additional search terms to the query. For example, when a user executed a hybrid Promote search, ViewFinder:

1. moved the keyframe of the promoted result to the middle panel of the client for visual reference
2. retrieved significant[78] keywords indexed for the promoted video and gathered terms entered by the user
3. retrieved video segments (results) with matching keywords[79]
4. retrieved keyframes with similar visual characteristics

---

[78] Significant keywords have a *tf•idf* weight that exceeds a predetermined threshold.

[79] The *Promote* keyword function performs an ‘OR’ search. In the experimental version of ViewFinder, segments with two or more matching keywords have *tf·idf* values combined and an overall relevancy score for that particular shot is calculated.

5. normalized keyword and visual similarity values using a 0.0 to 1.0 relevancy scale; the higher the score the more relevant search result

6. computed similarity scores for visual and keyword values using weighting inputs by the user (see search panel description below)

7. computed overall similarity score for returned results

8. sorted and return results to user, or client

The Details feature retrieved and presented additional information about a selected search result. The metadata displayed by the Details function varied by (ViewFinder) system. The experimental version of the ViewFinder displayed the video title, transcript terms, IDs, and a video abstract for each segment. Video details are displayed in a separate pop-up window.

The search panel, on the right-hand side of the interface, contained several other search and browse features. A keyword search was implemented to search ASR outputs, or video transcripts. For each of the systems presented in Table 5, if a query contained more than one keyword, the system performed an 'OR' search by default. However, unlike the TRECVID 2004 system, users of the NASA experimental system were able to formulate and execute 'AND' queries. The overall relevancy computation for each returned result included the simple scalar product of *tf•idf* weightings for all matching terms.

ViewFinder also provided several browsing options. The browsing feature is selected from the menu at the top (right) of searching panel. The browsing options also vary by ViewFinder version. The NASA experimental version of ViewFinder allowed users to browse by series title and duration. The TRECVID 2004 system (Table 5) allowed users to browse by date, source, and the combination of date and source. Once a user selected a browsing option, the matching set of choices are retrieved and listed in the box below the menu. Users could then select one of the choices and click Search, where results are retrieved and displayed.

The More Clips button, Back button, and feedback field are additional features of the search panel. The More Clips and Back button are used to review the returned results. The feedback field is used for information display only. The feedback field lists the last performed search query, number of results, number of results that match by different visual characteristics, and number of results that match by keyword.

Chapter 6

Results

The experimental data of this study was primarily collected and analyzed using quantitative techniques; however, several qualitative methods were also employed. In addition, both objective and subjective analyses were performed as part of this study. Objective results depict a variety of experimental parameters involving user demographics, task performance, and actions (or interactions) performed by the subjects. Subjective data, on the other hand, describes the subjects' perceptions about system effectiveness, interface intuitiveness, and task completion and representation. Results from both objective and subjective analyses were used to identify relationships and differences among these experimental parameters and corresponding factors. A discussion of experimental results follows.

6.1 Subject Sample and Demographics

Twenty-eight (28) subjects were recruited and successfully completed one full search experiment, apiece. As previously described, subject demographics were collected using the pre-experiment questionnaire, located in Appendix N. Demographic information collected for this study pertained to gender, age, level(s) of education, academic major(s), student status, current occupation, and grade levels taught or interested[80] in teaching.

From examining Tables 6 through 12, readers can begin to observe distributions and sampling among the subject pool. It was required that each of the subjects recruited for this study should have taught in an area of Science Education, or had some schooling or formal training for a future teaching position in science. All 28 subjects met this requirement. Tables 6 through 12 depict the subject pool according to different sub-samples.

First, readers can observe how the subject sample was distributed by gender. Table 6 presents that both male and female science educators were recruited and participated in this study; however, it also shows that there more females than males in the sample. A total of 18 females, or 64.3% of the participants, comprised the sample while 35.7%, 10 subjects, were male.

Table 6: Subject sample by gender.

Gender	Frequency	Percent
Male	10	35.7
Female	18	64.3
Total	28	100.0

[80]Teaching interests apply to the fulltime students recruited for this study.

Table 7 presents the sample by certain age groups. The sampling of subjects by age shows a satisfactory distribution among the participants. The overall range in age was 35 years, with a minimum of 21 and a maximum of 56. The age groups (presented in Table 7) were broken into ten year intervals, including groups for those in their twenties, thirties, forties, and fifties. This distribution shows more subjects in the earlier age groups, including a total of 13 out of 28 (46.4%) in their twenties and 5 out of 28 (17.9%) in their thirties. The older groups were also well represented with seven subjects, 25% of the sample, in the forties group and three, 10.7%, in the fifties group. Table 7 also shows that the overall mean *(M=)* age of this sample was 35.18 years with a standard deviation *(SD=)* of 11.37.

Table 7: Subject sample by age groups.

Age Groups	**Frequency**	**Percent**
20 – 29	13	46.4
30 – 39	5	17.9
40 – 49	7	25.0
50 – 59	3	10.7
Total	28	100.0
Overall Mean (SD)	35.18 *(11.37)*	

This study also collected the highest level of education achieved by the subjects. Because it's mandatory for science educators, or school teachers, to hold four-year degrees from a university, all subjects from this study either possessed a Bachelor's degree or anticipated graduating from a four-year program within the 2006 – 2007 academic year. In addition, many of the subjects had completed degrees beyond a Bachelor's. Table 8 shows that a total of 17, or 60.7% of the sample, held a Master's degree or higher[81]. As for the other 11 subjects in the sample, a Bachelor's degree was the highest level of education, either already possessed or anticipated.

Table 8: Subject sample by highest level of education achieved.

Highest Degree	Frequency	Percent
Bachelors	11	39.3
Masters	17	60.7
Total	28	100.0

Table 9 presents the academic majors of all subjects. Considering the breadth of Science Education, a wide variety of backgrounds within education and science was anticipated among the participants of this study. The academic backgrounds of the subjects ranged from elementary education to chemical engineering. Elementary education, a more conventional and standardized degree program, was the most common individual major at 42.9%, or 12 subjects. Secondary and high-school education degrees, on the other hand, typically entail more specialized areas of study; thus, the academic backgrounds of high-school and secondary teachers were more diverse. For example, subjects who taught secondary or high-school education held degrees in biology, chemistry, chemical engineering, earth science, general science, geography, library science, and mathematics education.

As previously mentioned, some of the participants were currently working toward and anticipating a degree in Science Education within the academic year. The researchers of this study recruited the full-time Science Education students from the School of Education at Indiana University, Bloomington (see Appendix B for recruiting

[81] None of the participants held a doctorate.

script). In addition, the experimental sample contained subjects who were already in full-time teaching positions while also working towards graduate credits. Table 10 shows the number of subjects who were students at the time of the experiments, including those in full-time undergraduate programs and those enrolled part-time in graduate credits. A total of 11 (39.3%) subjects were considered to be current students; seventeen, or 60.7%, of the subjects were not enrolled in any college credit whatsoever.

Table 9: Subject sample by academic major.

Major(s)	Frequency	Percent
Biology	1	3.6
Chemistry, Chemical Education, Chemical Engineering	3	10.7
Earth Science Education	1	3.6
Science Education	8	28.6
Elementary Education	12	42.9
Geography	1	3.6
Library Science / Education	1	3.6
Mathematics Education	1	3.6
Total	28	100.0

Table 10: Subject sample by current student status.

Current Students	Frequency	Percent
Yes	11	39.3
No	17	60.7
Total	28	100.0

The subject demographics collected for this study also included the current occupations of the participants. Table 11 shows a total of four different positions held by the subjects. These positions ranged from a current school administrator, teachers, a school librarian, and full-time students. The largest sub-sample of participants, according to occupation, was full-time teachers with a total of 21, or 75% of the sample. Full-time students comprised the second largest sub-sample with 5 out of 28 (17.9%) of the subjects. There was also one, 3.6%, school administrator and one librarian in the sample[82].

Table 11: Subject sample by current occupation.

Current Occupation or Position	Frequency	Percent
School Administrator, Former Science Teacher	1	3.6
Student	5	17.9
Teacher	21	75.0
Librarian, Former Science Teacher	1	3.6
Total	28	100.0

The grade levels taught, or planned to teach, by the subjects was another demographic collected for this study (Table 12). The distribution of the different grade levels is satisfactory. The grades taught included elementary, secondary (middle-school), and high-school. When analyzing the sample by grade levels, elementary teachers comprised the largest sub-sample with a total of 14 out of 28, or 50%, of the subjects. Meanwhile, high-school and secondary educators comprised the other 50% of the sample with 9 (32.1%) and 5 (17.9%) subjects, respectively.

Table 12: Subject sample by primary grade level taught, or planning to teach.

Levels Taught	Frequency	Percent
Elementary	14	50.0
Secondary	5	17.9
High School	9	32.1
Total	28	100.0

[82] The school administrator and librarian that participated in this study were former science teachers.

Also, as part of the pre-experiment questionnaire, subjects' familiarity with various technologies and other educational-related activities were collected and analyzed. Tables 13 through 16 summarize these findings by reporting means and standard deviations. The pre-experiment questions assessing technological and educational skills were based on five-point scales, where zero indicated low familiarity and five indicated high familiarity. The technological skills examined for this study included subjects' familiarity with point and click interfaces[83], and searching on-line databases systems, WWW search engines, video retrieval systems, and other educational research systems. The pre-experiment questionnaire also collected data about prior experiences and familiarity with other non-technical skills, including preparing Science Education projects and lessons plans, and incorporating video into classroom activities.

Table 13: Pre-experiment results by age groups.

Age Group *(SD)*	**Point and Click Interface *(SD)***	**Searching Online Systems/ DBs *(SD)***	**Searching WWW Search Engines *(SD)***	**Searching Video Retrieval Systems *(SD)***	**Searching Other Educational Systems *(SD)***	**Preparing Science Ed Projects, Lessons *(SD)***	**Using Video In Classroom & Assignments *(SD)***
20 – 29	4.77 *(0.60)*	3.31 *(0.63)*	4.69 *(0.48)*	2.31 *(1.11)*	1.85 *(1.28)*	3.69 *(0.86)*	2.92 *(1.04)*
30 – 39	4.40 *(0.89)*	2.40 *(1.14)*	4.80 *(0.45)*	2.60 *(0.89)*	2.00 *(1.41)*	4.20 *(0.84)*	3.20 *(1.10)*
40 – 49	4.43 *(0.79)*	2.86 *(1.35)*	4.14 *(1.07)*	2.14 *(0.90)*	2.14 *(1.46)*	4.00 *(1.16)*	2.71 *(1.25)*
50 – 59	4.50 *(0.71)*	1.67 *(1.16)*	3.00 *(0.00)*	2.33 *(2.31)*	1.00 *(0.00)*	3.67 *(1.53)*	3.67 *(2.31)*
Total	4.59 *(0.69)*	2.86 *(1.08)*	4.44 *(0.80)*	2.32 *(1.12)*	1.85 *(1.26)*	3.86 *(0.97)*	3.00 *(1.22)*

By examining the rows marked *Total* in Tables 13 through 16, readers can review the overall scores for each of the pre-experiment questions. The technological skill that scored the highest, in regards to familiarity, was using point and click interfaces at an

[83] Examples of point and click interfaces, as presented in the pre-experiment questionnaire, include Windows and Mac operating systems.

overall mean of 4.59 out of 5.0. Searching on-line Web search engines ranked second highest with a mean of 4.44. Subjects were shown to be significantly less familiar, at 0.01, with searching on-line commercial database systems, such as LexisNexis and EBSCO, with a mean of 2.86, and using video retrieval systems, or video digital libraries, at 2.32. The subjects were least familiar with naming and judging familiarity with other educational retrieval systems at a mean score of 1.85 out of 5.0. Refer to Appendix N to view the pre-test questionnaire.

From examining results from the pre-experiment questionnaire, the researchers were also able to assess subjects' experience with some of the more general tasks facing K – 12 science educators. For example, subjects demonstrated a moderate to high familiarity with preparing science educational projects and/or lessons plans at a mean score of 3.86 out of 5.0, apiece. In addition, the overall mean for using and incorporating video into classroom activities was measured at 3.0 out of 5.0.

Pre-experiment analysis was also used to examine subjects' familiarity with certain technologies and educational-related activities across different sub-samples. The technological and educational skills, as just described, were evaluated according to different age groups, education levels (possessed by the subjects), grade level(s) taught, and current occupation. First, there were several distinctions among the technological skills of the different age groups. For example, subjects' familiarity with using Web search engines showed that younger subjects, i.e. those in their twenties and thirties, demonstrated higher levels than the older sub-samples[84]. In addition, there were significant differences[85] between subjects in their twenties and all other age groups when examining familiarity with searching commercial database systems; the twenties group scored a mean of 3.31, while the other groups ranged from 1.67 for the fifties group to 2.86 for the forties group.

There were not as many distinctions among the age groups when comparing experience with other technological and educational skills, as assessed on the pre-experiment questionnaire. The overall mean of using point and click interfaces was quite high across all age groups, where the range of means was only 0.37[86]. Similarly, when comparing the different age groups based on subjects' familiarity with searching video systems and preparing science educational lessons, the overall ranges were 0.46 and 0.53, respectively. A slightly higher degree of difference was observed when comparing age groups by their experience with incorporating video into classroom activities, which produced an overall mean range of 0.96[87].

The next demographic used for comparing technological and educational skills across the sample included the level(s) of education achieved by the subjects.

---

[84] When asked about their familiarity with using Web search engines, the twenties and thirties age groups scored means of 4.69 and 4.80, respectively, while the forties and fifties groups achieved mean scores of 4.14 and 3.00.

[85] At 0.05 significance level.

[86] Subjects in their twenties showed the highest familiarity with a mean of 4.77, while subjects in the thirties group scored the lowest with an overall mean of 4.40

[87] The oldest age group, i.e. subjects in their fifties, scored a mean of 3.67 while the forties group scored an experimental low of 2.71.

Demographic data showed (Table 14) that subjects belonged to one of two groups. This (educational level) sub-sample included subjects who possessed or working toward[88] a Bachelor's degree and those who held a Master's degree or higher[89]. One trend that was observed from this analysis included that subjects whose highest level of education was a Bachelor's degree showed greater familiarity with the majority of these technological skills. For example, subjects with a Bachelor's degree exhibited marginally higher levels of familiarity with using point and click interfaces, and searching Web search engines and video retrieval systems. In addition, those with a Bachelor's degree also had significantly more experience with searching on-line commercial databases, at a mean of 3.27, compared to those with a Master's or higher, at 2.59. Those with a Master's degree did score slightly higher with naming and searching other educational-based research systems[90].

Table 14: Pre-experiment results by level of education achieved.

Highest Level of Education Achieved	**Point and Click Interface *(SD)***	**Searching Online Systems/ DBs *(SD)***	**Searching WWW Search Engines *(SD)***	**Searching Video Retrieval Systems *(SD)***	**Searching Other Educational Systems *(SD)***	**Preparing Science Ed Projects, Lessons *(SD)***	**Using Video In Classroom & Assignments *(SD)***
Bachelors	4.73 *(0.65)*	3.27 *(0.65)*	4.73 *(0.47)*	2.36 *(1.21)*	1.82 *(1.33)*	3.45 *(0.82)*	2.73 *(0.91)*
Masters	4.50 *(0.73)*	2.59 *(1.23)*	4.25 *(0.93)*	2.29 *(1.11)*	1.87 *(1.26)*	4.12 *(0.99)*	3.18 *(1.38)*
Total	4.59 *(0.69)*	2.86 *(1.08)*	4.44 *(0.80)*	2.32 *(1.12)*	1.85 *(1.26)*	3.86 *(0.97)*	3.00 *(1.22)*

88 The full-time Science Education students all anticipated degree conferral within the 2006 – 2007 academic year.

89 It is important to note that although some of the participants did possess degree(s), certificate(s) or credit hours beyond a Master's, none actually held a doctorate in their field.

90 This comparison was not statistically significant.

Another trend that was observed after analyzing these skills across different levels of education included that subjects with a Master's degree (or higher) had more experience with performing these educational-related tasks. Moreover, subjects with a Master's degree scored means of 4.12 and 3.18 (out of 5.0) when estimating their familiarity with preparing Science Education lessons plans and incorporating video into classroom activities. Meanwhile, subjects with a Bachelor's degree as their highest level of education averaged 3.45 and 2.73 for their experience with these two skills.

This study also assessed technological and education-related skills across the different grade levels taught[91] by the subjects. As previously described, subjects fell into one of three different categories, including elementary, secondary, or high school teachers. Again, across all sub-samples, subjects were shown to be most familiar with using point and click interfaces and searching Web search engines, and least familiar with searching video systems and other educational research systems. Exact scores and standard deviations are listed in Table 15.

Table 15: Pre-experiment results by grade taught, or will teach.

Grades Taught or Interested in Teaching	**Point and Click Interface *(SD)***	**Searching Online Systems/ DBs *(SD)***	**Searching WWW Search Engines *(SD)***	**Searching Video Retrieval Systems *(SD)***	**Searching Other Educational Systems *(SD)***	**Preparing Science Ed Projects, Lessons *(SD)***	**Using Video In Classroom & Assignments *(SD)***
Elementary	4.54 *(0.78)*	2.50 *(1.02)*	4.38 *(0.77)*	2.29 *(1.14)*	1.46 *(0.78)*	3.71 *(0.91)*	3.14 *(1.03)*
Secondary	4.20 *(0.84)*	3.20 *(1.30)*	3.80 *(1.10)*	2.00 *(1.00)*	1.80 *(1.30)*	3.80 *(1.30)*	2.40 *(1.52)*
High School	4.89 *(0.33)*	3.22 *(0.97)*	4.89 *(0.33)*	2.56 *(1.24)*	2.44 *(1.67)*	4.11 *(0.93)*	3.11 *(1.36)*
Total	4.59 *(0.69)*	2.86 *(1.08)*	4.44 *(0.80)*	2.32 *(1.12)*	1.85 *(1.26)*	3.86 *(0.97)*	3.00 *(1.22)*

[91] Currently, formerly, or in the future.

One important trend that was observed from Table 15 included that subjects who taught high school ranked highest (in regards to prior experience) across all technological skills. High school teachers scored a mean of 4.89 for using point and click interfaces, 3.22 for searching on-line commercial database systems, 4.89 for searching the Web, 2.56 for using video retrieval systems, and 2.44 for searching other educational systems. These scores were significantly higher than many of the low averages for each of these factors, including secondary teachers' familiarity with point click interfaces (4.20), Web search engines (3.80), and video retrieval systems (2.00), and elementary teachers' experience with searching on-line commercial databases (2.50) and other educational systems (1.46).

The range of means weren't as vast when comparing subjects' familiarity with the educational-related activities across different grade levels taught. While high school teachers showed the highest familiarity with preparing Science Education projects and lesson plans, with a mean score of 4.11, elementary teachers scored the lowest at 3.71, which resulted in an overall range of means at only 0.40. Measuring subjects' experience with incorporating video into the classroom demonstrated that elementary teachers ranked the highest, at an average of 3.14, compared to high school and secondary teachers at scores of 3.11 and 2.40, respectively[92].

Lastly, in regards to demographics, this study examined how the subjects' technological and educational skills varied according to their current occupations. As shown in Table 16, there were four different occupations currently held by the subjects in

[92] An overall range of 0.74.

the sample. The current occupations of the subjects included teachers, full-time students, one librarian, and one school administrator[93].

Table 16: Pre-experiment results by current occupation.

Current Occupation	Point and Click Interface *(SD)*	Searching Online Systems/ DBs *(SD)*	Searching WWW Search Engines *(SD)*	Searching Video Retrieval Systems *(SD)*	Searching Other Educational Systems *(SD)*	Preparing Science Ed Projects, Lessons *(SD)*	Using Video In Classroom & Assignments *(SD)*
Admin.[94]	5.00 *(.)*	2.00 *(.)*	5.00 *(.)*	3.00 *(.)*	2.00 *(.)*	4.00 *(.)*	3.00 *(.)*
Student	5.00 *(0.00)*	3.40 *(0.55)*	4.80 *(0.45)*	2.40 *(1.67)*	1.80 *(1.79)*	3.60 *(0.89)*	2.20 *(1.10)*
Teacher	4.45 *(0.76)*	2.67 *(1.07)*	4.30 *(0.87)*	2.24 *(1.04)*	1.70 *(0.98)*	3.90 *(1.04)*	3.10 *(1.18)*
Librarian[4]	5.00 *(.)*	5.00 *(.)*	5.00 *(.)*	3.00 *(.)*	5.00 *(.)*	4.00 *(.)*	5.00 *(.)*
Total	4.59 *(0.69)*	2.86 *(1.08)*	4.44 *(0.80)*	2.32 *(1.12)*	1.85 *(1.26)*	3.86 *(0.97)*	3.00 *(1.22)*

Results were analyzed across these different sub-samples, and variances among full-time students and current teachers[95] are reported. Overall, full-time students possessed the highest levels of familiarity with all of these technical skills. Full-time students scored an overall mean of 5.0 out of 5.0 for using point and click interfaces, and 4.80 for searching on-line Web search engines. These results are compared to current teachers who averaged 4.45 for point click interfaces and 4.30 for searching the Web. The student sub-sample also demonstrated higher familiarity with searching on-line commercial databases, video retrieval systems, and other educational systems with mean

[93] Considering the sample contained only one school administrator and one librarian, these occupations was not subjected to more comparison with the other occupations. These scores, however, are reported in Table 11.

[94] The school administrator and school librarian recruited for this study were both former science teachers.

[95] The teachers used in this analysis included those from all grade levels, i.e. elementary, secondary, and high school.

scores of 3.40, 2.40, and 1.80, respectively. This is compared to 2.67, 2.24, and 1.70, the mean scores of current teachers. However, results showed that teachers exhibited greater experience with the other educational-related activities. For example, teachers scored 3.90 and students achieved 3.60 when evaluating familiarity with preparing Science Education lessons. In addition, mean scores of 3.10 (teachers) and 2.20 (students) were observed when assessing their experience with incorporating video into classroom activities.

6.2 Objective Results

"During-test" measures comprised the objective results of this study. During-test results were collected by monitoring each subject perform the experimental search tasks and examining the server logs that were accumulated during each experiment. The collection of during-test results included recording and assessing task completion ratios, start and end times, errors, and the use of each interface feature and function. The experimental variables and factors corresponding to the objective results were analyzed using a variety of quantitative methods.

6.2.1 During-Test Results

One set of parameters analyzed for this study were included in the "during-test" results. During-test results were collected to assess subjects' performance on each of the search tasks. For this particular study, performance was measured using task completion ratios, completion times, and the number of steps and errors[96]. In addition, the actions (or interactions) performed by the subjects, including the use of each interface feature and function, were also recorded and analyzed. These during-test results were measured and compared across all systems variants and task types. During-test results are summarized in Tables 17 through 22.

6.2.1.1 Task Performance Results

By first examining Table 17, readers can observe the composite means for task performances, including all completion ratios, times, steps, and errors. Also, in Table 17, the row marked *Overall* presents the composite means and standard deviations (*SD*) for task performance measures across all system variants and task types. From examining these results, readers can be observe that the overall task completion ratio, for the entire experiment, was 75% with a *SD*=37.91 percentage points. The composite results also indicated that the average time spent performing each task was 2:51 with a *SD*=1.63 minutes. It's also important to look at the composite means for the number of steps and errors, which were assessed at 8.30 and 2.14 per search task, respectively.

Table 17 also presents the collective means achieved for different task types. From these results, readers can compare task performances on textual, visual, and hybrid tasks. Results showed that visual-only tasks demonstrated higher scores of task completion with a mean of 85.71% and *SD*=31.79, compared to 76.19% for text-only and 69.05% for hybrid tasks. In addition, visual tasks produced lower task times at an overall

[96] Errors were as unmistakable incorrect queries or misuse of any interface feature or function.

mean of 2:31, while times for textual tasks were slightly higher at 2:40 and hybrid tasks averaged 3:07. However, text-only tasks exhibited fewer steps (mean of 6.14) and errors (mean of 1.64) overall. Both visual and hybrid tasks produced higher averages for steps and errors, at mean scores of 7.83 and 9.61 (steps) and 1.98 and 2.46 (errors).

Table 17: Composite task performance results.

Task Type	Task Completion *(SD)*	Time of completion (*SD*)	Steps *(SD)*	Errors *(SD)*
Overall	75% *(37.91)*	2:51 *(1.63)*	8.30 *(6.25)*	2.14 *(2.50)*
Text-only	76.19% *(40.18)*	2:40 *(1.87)*	6.14 *(6.39)*	1.64 *(2.80)*
Visual-only	85.71% *(31.79)*	2:31 *(1.55)*	7.83 *(5.23)*	1.98 *(2.27)*
Hybrid	69.05% *(38.73)*	3:07 *(1.52)*	9.61 *(6.38)*	2.46 *(2.43)*
Easy	80.95% *(39.74)*	1:34 *(1.19)*	5.31 *(6.39)*	1.55 *(2.91)*
Easy-Textual	80.95% *(40.24)*	1:26 *(1.35)*	4.71 *(8.33)*	1.67 *(3.86)*
Easy-Visual	80.95% *(40.24)*	1:43 *(1.01)*	5.90 *(3.69)*	1.43 *(1.54)*
Complex	80.95% *(33.04)*	3:37 *(1.53)*	8.67 *(4.81)*	2.07 *(2.11)*
Complex-Textual	71.43% *(40.53)*	3:54 *(1.45)*	7.57 *(3.19)*	1.62 *(1.07)*
Complex-Visual	90.48% *(20.12)*	3:20 *(1.59)*	9.76 *(5.89)*	2.52 *(2.75)*
Combination	73.81% *(40.18)*	2:52 *(1.58)*	8.36 *(6.50)*	2.10 *(2.54)*
Combo-Complex	64.29% *(37.10)*	3:21 *(1.43)*	10.86 *(6.08)*	2.83 *(2.28)*

Other trends can be observed from examining means from the various task categories, including differences between easy, complex, combination, and combo-complex tasks. According to these results, complex-visual tasks achieved the highest ratio of task completion at 90.48% and *SD*=20.12. Complex and easy tasks each averaged an overall score of 80.95% for task completion. Combination tasks exhibited an overall task completion ratio at 73.81% while combo-complex tasks scored lower among these individual task types at 64.29%.

Similar comparisons can be made using task completion times. When assessing completion times across different task types, easy-textual tasks exhibited a lower mean at 1:26 per task, while complex-textual tasks required 3:54. Readers can observe other completion times achieved through this analysis, including easy tasks averaging 1:34, combination tasks at 2:52, combo-complex tasks at 3:21, and complex tasks at 3:37.

Analyzing the number of steps and errors performed for each task type produced somewhat similar trends across all system variants. Easy-textual tasks required fewer steps on average at 4.71 per task while combo-complex tasks exhibited the most steps at a mean of 10.86. Overall, easy tasks required the fewest number of steps at an average of 5.31 per task, which can be compared to combination tasks at 8.36, complex tasks at 8.67, and combo-complex tasks with an experimental high of 10.86. Analyzing the number of errors also showed easy tasks to produce a lower mean score at 1.55 per task, while complex tasks demonstrated an average of 2.07, combination tasks at 2.10, and combo-complex tasks at 2.83 per task.

Next, readers can observe these task performance measures across different system variants[97] (Table 18). Composite performance results, including task completion ratios, completion times, and the number of steps and errors, are reported for each system variant in the rows marked *Overall*. These scores demonstrated that System 3, i.e. the full system, produced the highest percentage of task completion at a mean of 89.29% and SD=24.71. The mean for System 3 was followed by the results for System 1, which exhibited a task completion ratio of 83.93%. System variant 2, on the other hand, scored a significantly lower task completion ratio than the other system variants at an average of 51.79%. Task completion times also varied, and exhibited that System variant 3 resulted in the least amount of time at 2:21, followed by System 1 at 2:46 and System 2 at 3:26. The average number of steps and errors showed similar patterns where System 3

[97] See page 74 for description of each system variant.

demonstrated fewer steps (*M*=7.27) and errors (*M*=1.63). Again, System variant 3 was followed by System 1, which averaged 8.05 steps and 2.00 errors per task. System 2 demonstrated higher averages for both steps and errors at scores of 9.57 and 2.79, respectively.

Table 18: Task performance results by system variant and task type.

System	Task Type	Task Completion *(SD)*	Time of completion (*SD*)	Steps *(SD)*	Errors *(SD)*
1	**Overall**	83.93% *(34.52)*	2:46 *(1.63)*	8.05 *(6.65)*	2.00 *(2.72)*
	Text-only	71.43% *(46.88)*	3:26 *(1.70)*	9.36 *(9.03)*	2.93 *(4.39)*
	Visual-only	85.71% *(36.61)*	2:00 *(1.24)*	6.21 *(4.53)*	1.21 *(1.48)*
	Hybrid	89.29% *(24.93)*	2:50 *(1.65)*	8.32 *(6.18)*	1.93 *(1.99)*
	Easy	57.14% *(51.36)*	2:17 *(1.44)*	8.50 *(9.25)*	3.00 *(4.39)*
	Easy-Textual	42.86% *(53.45)*	2:43 *(1.60)*	10.86 *(12.59)*	4.57 *(5.86)*
	Easy-Visual	71.43% *(48.76)*	1:51 *(1.21)*	6.14 *(3.72)*	1.43 *(1.27)*
	Complex	100% *(0.00)*	3:09 *(1.75)*	7.07 *(4.55)*	1.14 *(1.41)*
	Complex-Textual	100% *(0.00)*	4:09 *(1.57)*	7.86 *(3.58)*	1.29 *(1.11)*
	Complex Visual	100% *(0.00)*	2:09 *(1.35)*	6.29 *(5.53)*	1.00 *(1.73)*
	Combination	92.86% *(26.73)*	2:15 *(1.63)*	5.71 *(5.82)*	1.14 *(1.70)*
	Combo-Complex	85.71% *(23.44)*	3:26 *(1.50)*	10.93 *(5.54)*	2.71 *(2.02)*
2	**Overall**	51.79% *(41.52)*	3:26 *(1.67)*	9.57 *(6.96)*	2.79 *(2.88)*
	Text-only	64.29% *(45.69)*	2:32 *(2.12)*	5.07 *(4.71)*	1.07 *(1.21)*
	Visual-only	85.71% *(23.44)*	3:09 *(1.88)*	8.00 *(5.92)*	2.57 *(3.30)*
	Hybrid	28.57% *(31.71)*	4:02 *(1.00)*	12.61 *(7.06)*	3.75 *(2.91)*
	Easy	100% *(0.00)*	1:11 *(0.95)*	2.36 *(2.68)*	0.29 *(0.61)*
	Easy-Textual	100% *(0.00)*	0:39 *(24.40)*	1.00 *(0.00)*	0.00 *(0.00)*
	Easy-Visual	100% *(0.00)*	1:43 *(1.11)*	3.71 *(3.35)*	0.57 *(0.79)*
	Complex	50.00% *(39.22)*	4:30 *(1.16)*	10.71 *(4.14)*	3.36 *(2.85)*
	Complex-Textual	28.57% *(39.34)*	4:26 *(1.13)*	9.14 *(3.08)*	2.14 *(0.69)*
	Complex-Visual	71.43% *(26.73)*	4:34 *(1.27)*	12.29 *(4.68)*	4.57 *(3.69)*
	Combination	32.14% *(37.25)*	3:56 *(1.07)*	11.93 *(7.46)*	3.29 *(3.38)*
	Combo-Complex	25% *(25.94)*	4:09 *(0.95)*	13.29 *(6.84)*	4.21 *(2.39)*
3	**Overall**	89.29% *(24.71)*	2:21 *(1.42)*	7.27 *(4.82)*	1.63 *(1.58)*
	Text-only	92.86% *(18.16)*	2:02 *(1.60)*	4.00 *(2.80)*	0.93 *(1.14)*
	Visual-only	85.71% *(36.31)*	2:26 *(1.34)*	9.29 *(5.06)*	2.14 *(1.46)*
	Hybrid	89.29% *(20.89)*	2:28 *(1.40)*	7.89 *(4.81)*	1.71 *(1.74)*
	Easy	85.71% *(36.31)*	1:15 *(0.80)*	5.07 *(3.95)*	1.36 *(1.74)*
	Easy-Textual	100% *(0.00)*	0:56 *(0.73)*	2.29 *(2.36)*	0.43 *(0.79)*
	Easy-Visual	71.43% *(48.80)*	1:34 *(0.79)*	7.86 *(3.19)*	2.29 *(1.98)*
	Complex	92.86% *(18.16)*	3:13 *(1.31)*	8.21 *(5.25)*	1.71 *(1.07)*
	Complex-Textual	85.71% *(24.40)*	3:09 *(1.46)*	5.71 *(2.14)*	1.43 *(1.27)*
	Complex-Visual	100% *(0.00)*	3:17 *(1.25)*	10.71 *(6.37)*	2.00 *(0.82)*
	Combination	96.43% *(13.36)*	2:26 *(1.50)*	7.43 *(4.64)*	1.86 *(1.83)*
	Combo-Complex	82.14% *(24.86)*	2:30 *(1.34)*	8.36 *(5.11)*	1.57 *(1.70)*

This analysis also allows a closer examination of the variances between task type(s) and individual system variants, collectively. Examining the during-test results for System 1 showed certain differences among textual, visual, and hybrid search tasks. These results presented that hybrid search tasks produced a task completion ratio of 89.29%, which was followed by visual-only tasks at 85.71% and textual tasks at 71.43%. However, when using System variant 1, visual-only tasks required a lower completion time on average at 2:00, along with a fewer number of steps and errors at 6.21 and 1.21 per task, respectively. These results were significantly different from the averages for textual tasks, including 3:26 for task completion, and 9.36 steps and 2.93 errors.

When examining the during-test results for System 1 across easy, complex, combination, and combo-complex search tasks, complex tasks showed a higher task completion ratio at 100%. Subsequently, all complex-textual and complex-visual tasks were answered successfully across the entire experiment when using System 1. Combination tasks closely followed with a 92.86% completion ratio, and combo-complex task also scored relatively high at 85.71%. Easy tasks scored significantly lower than complex, combination, and combo-complex tasks with an average completion ratio of 57.14%. Analyzing completion times and the numbers of steps and errors, when evaluating System variant 1, demonstrated that combination tasks again achieved positive scores. Moreover, combination tasks showed a mean completion time of 2:15 and an average number of steps at 5.71 and errors at 1.14 per task. Scores for combination tasks differed from combo-complex tasks, which required a mean time of 3:26 and an average of 10.93 steps. Easy tasks exhibited an overall error rate of 3.00 for System 1, but easy-textual tasks resulted in more errors and steps at 4.57 and 10.86, apiece.

Composite task performance and during-test results for System variant 2 revealed several other distinctions. The overall task completion ratio for System 2 was 51.79%, while the mean task completion time was 3:26, and the numbers of steps and errors averaged 9.57 and 2.79, respectively.

When evaluating System 2 across textual, visual, and hybrid task types, there were other differences among task performances and during-test results. Visual tasks resulted in a higher percentage of task completion for System 2 at 85.71%, followed by textual tasks at 64.29% and hybrid tasks at 28.57%. Hybrid tasks also demonstrated some lower averages for completion times, steps, and errors. Moreover, 4:02, 12.61 steps, and 3.75 errors were the averages reported for each hybrid task. Textual tasks, on the other hand, demonstrated more positive completion times at 2:32 per task, and also satisfactory means for steps (5.07) and errors (1.07). Visual tasks scored a mean of 3:09 for completion time, and 8.00 and 2.57 for steps and errors, respectively.

Other categories of search tasks were used to evaluate System variant 2. All easy tasks, including easy-visual and easy-textual tasks, scored a 100% task completion ratio when using System 2. Complex tasks, however, produced a significantly lower task completion score at 50%, followed by combination tasks at 32.14%, and combo-complex tasks at only 25.0%. In addition, easy tasks required less time while using System 2 with a mean of 1:11 per task, which also included an experimental low of 0:39 for easy-textual tasks. Readers can compare the task completion times for easy tasks with the 4:30 for complex tasks, 4:09 for combo-complex tasks, and 3:56 seconds for combination tasks, which are all significantly different. Analyzing the number of steps and errors for the easy tasks again showed positive results. Moreover, easy tasks averaged 2.36 steps and only 0.29 errors per task[98]. Compare these scores to the 13.29 steps averaged for the combo-complex tasks, 11.93 steps for combination tasks, and 10.71 for the complex tasks. Errors rates for these more-ineffective search tasks, while evaluating System 2, included combo-complex tasks at a mean of 4.21 per tasks, complex task at 3.36, and combination tasks at 3.29.

Lastly, it was important to examine task performance scores for System variant 3. From looking at Table 18, readers can observe that System 3 exhibited better overall scores than the other system variants, particularly for task completion, completion times, and number of steps and errors[99].

From examining differences across textual, visual, and hybrid tasks, as they scored for System 3, each type achieved relatively similar levels for task performance. Textual tasks demonstrated a positive task completion score of 92.86% while visual tasks also scored a satisfactory mean at 85.71%. In regards to completion times, textual tasks produced a mean time of 2:02, followed by visual tasks at 2:26 and hybrid tasks at 2:28. Results were more mixed when analyzing the average number of steps and errors. Visual tasks produced the highest means for both steps and errors at 9.29 and 2.14, respectively. Hybrid tasks averaged 7.89 steps and 1.71 errors for each task. The task type that exhibited more positive results was textual tasks, which averaged 4.0 steps and 0.93 errors.

---

[98] These results are supported by the means for easy-textual tasks, which only averaged one (1.0) step per task, and there were no (0.00) errors across the entire experiment.

[99] System variant 3 scored an overall mean of 89.29% (*SD*=24.71) for tasks completed, and also averaged 2:21 (*SD*=1.42) for task completion time, 7.27 (*SD*=4.82) steps per task, and 1.63 (*SD*=1.58) errors per task.

Results also revealed differences for System variant 3 among easy, complex, combination, and combo-complex tasks. Combination tasks scored a higher overall percentage of task completion at 96.43%. Results for complex tasks were next highest averaging 92.86%, along with easy tasks at 85.71% and combo-complex tasks at 82.14%. Easy tasks produced positive results for completion times and the number of steps and errors. The evaluation of System 3 showed that easy tasks averaged only 1:15, 5.07 steps, and 1.36 errors per task. Combination tasks also demonstrated significant averages for completion time (2:26) and number of steps (7.43). Other less-productive scores included the mean completion time of complex tasks at 3:13 and the number of steps for combo-complex tasks, which averaged 8.36 per task.

6.2.1.2 Use of Interface Features and Functions

During-test results were also used to measure and evaluate the use of individual interface features and functions. The results generated through this analysis are presented in Tables 19 through 22. These tables present the means (*M*) and standard deviations (*SD*) for the number of times each search feature (or technique) was used per search task. These results are important because they exhibit trends among the use of interfaces features and functions for different task types and system variants.

Table 19 presents composite results for certain interface features and functions across different system variants and task categories. By examining the row marked *Overall* (Table 19), readers can observe the average number of times each feature was used for all search tasks. These scores indicated that, across all task categories and system variants, keyword searching was employed the most at an average of 2.57 times per task (*SD*=1.56). The keyword search was closely followed by subjects' use of the Details feature, with an overall mean of 2.08 per task. Analysis also showed somewhat similar use of the Title Browse, results browse[100], and Promote features, with averages of 1.65, 1.82, and 1.20, respectively. The searching technique that exhibited a lower overall average included the reexamination of previously seen search results, or browsing back, at a mean of 0.43 per task.

Table 19: Use of interface features and functions, overall and among different task types.

Task Type	Keyword Searches *(SD)*	Browse Video Titles (*SD*)	Browse More Search Results *(SD)*	Reexamine Search Results *(SD)*	View Clip Details *(SD)*	Promote Search Result *(SD)*
Overall	2.57 *(1.56)*	1.65 *(1.94)*	1.82 *(2.85)*	0.43 *(0.96)*	2.08 *(2.89)*	1.20 *(1.78)*
Text-only	2.14 *(1.43)*	1.39 *(1.55)*	0.76 *(2.52)*	0.10 *(0.37)*	2.24 *(3.59)*	0.69 *(1.37)*
Visual-only	2.75 *(1.67)*	1.39 *(1.69)*	1.95 *(1.96)*	0.55 *(0.94)*	1.52 *(2.37)*	1.14 *(1.52)*
Hybrid	2.70 *(1.58)*	1.91 *(2.21)*	2.27 *(3.24)*	0.54 *(1.12)*	2.27 *(2.73)*	1.49 *(2.03)*
Easy	1.96 *(1.64)*	0.96 *(0.79)*	1.31 *(2.62)*	0.19 *(0.51)*	1.12 *(3.37)*	0.74 *(1.08)*
Easy-Textual	1.57 *(1.55)*	0.86 *(0.36)*	0.81 *(3.28)*	0.00 *(0.00)*	2.05 *(4.62)*	0.24 *(0.77)*
Easy-Visual	2.36 *(1.69)*	1.07 *(1.07)*	1.81 *(1.66)*	0.38 *(0.67)*	0.19 *(0.40)*	1.24 *(1.14)*
Complex	2.93 *(1.36)*	1.82 *(2.06)*	1.40 *(2.01)*	0.45 *(0.92)*	2.64 *(2.49)*	1.10 *(1.75)*
Complex-Textual	2.71 *(1.07)*	1.93 *(2.06)*	0.71 *(1.49)*	0.19 *(0.51)*	2.43 *(2.23)*	1.14 *(1.68)*
Complex-Visual	3.14 *(1.61)*	1.71 *(2.13)*	2.10 *(2.26)*	0.71 *(1.15)*	2.86 *(2.76)*	1.05 *(1.86)*
Combination	1.96 *(1.14)*	2.11 *(2.57)*	1.86 *(2.87)*	0.52 *(1.31)*	1.86 *(2.43)*	1.40 *(1.94)*
Combo-Complex	3.43 *(1.64)*	1.71 *(1.80)*	2.69 *(3.55)*	0.55 *(0.92)*	2.69 *(2.98)*	1.57 *(2.13)*

By further examining Table 19, readers can observe some differences and associations between the use of these interface features during textual, visual, and hybrid search tasks. This analysis of interface features presented some key observations, including how keyword searching was performed quite similarly across textual, visual, and hybrid task types. Moreover, keyword searches were shown to be employed an

[100] Search results browse includes using the More Clips feature of the user interface.

average of 2.75 times on visual-only tasks, 2.70 on hybrid tasks, and 2.14 on textual tasks. Results for the Title Browse feature demonstrated higher use during hybrid tasks, with a mean of 1.91, followed by textual and visual tasks, each at 1.39.

Another observation made from these results included differences among subjects' tendency to browse search results throughout textual, visual, and hybrid search tasks. Subjects were shown to browse additional search results an average of 2.27 times for hybrid tasks, while only browsing an average of 0.76 times on textual tasks. Scores depicting subjects' inclination to reexamine previously seen search results were low across each of these task categories. Results exhibited that subjects reexamined search results an average of 0.55 times per visual task and only 0.10 during textual tasks.

Similar patterns were also presented when analyzing subjects' use of the Details and Promote features. Results showed that subjects viewed (clip) details a mean of 2.27 times during hybrid tasks and a mean of 1.52 for visual tasks. Subjects also tended to employ Promote searches differently across textual, visual, and hybrid tasks. Overall, hybrid tasks showed higher levels of Promote searching at 1.49 per task, followed by a visual tasks at 1.14 and textual tasks at a mean of 0.69.

It was also important to explore the use of these interface features and functions across other categories of search tasks. Moreover, readers can observe which search features were used during easy, complex, combination, and combo-complex tasks. Overall, keyword searches were performed quite frequently for combo-complex tasks, at a mean of 3.43, while easy tasks demonstrated a lower average at 1.96. Further analysis of keyword searching revealed distinctions between easy-visual and easy-textual search tasks. These results showed that easy-visual tasks produced a higher rate of keyword searching (with a mean of 2.36) than easy-textual tasks (mean of 1.57). Differences among keyword searching were also apparent when comparing complex-textual with complex-visual tasks. Complex-visual tasks exhibited more frequent use of the keyword search as compared to complex-visual[101].

Results indicated other distinctions in how subjects used the Title Browse feature across different task types. For example, combination tasks demonstrated higher scores for the Title Browse at an overall mean of 2.11 per search task. Easy tasks, however, only averaged 0.96 for using the Title Browse. There were also no real differences in video browsing among the textual and visual forms of the evaluated task categories.

There were even clearer distinctions when analyzing subjects' tendency to browse search results. For example, combo-complex tasks required a relatively high amount of results browsing at an average of 2.69 per task, while easy tasks produced a lower score of 1.31. Also, results from this analysis indicated differences between the visual and textual forms of search tasks. Overall, easy-visual tasks averaged 1.81 (search result) browses per task, compared to easy-textual tasks at a mean of 0.81. This trend was also observed when analyzing complex tasks[102]. Moreover, complex-visual tasks exhibited significantly more results browsing than complex-textual tasks, at mean scores of 2.10 and 0.71, respectively.

Results depicting subjects' tendency to reexamine search results, throughout easy, complex, combination, and combo-complex tasks, were also generated and calculated.

[101] Complex-visual and complex-textual tasks averaged 3.14 and 2.71 keyword searches per task, respectively.

[102] As presented in Table 19, complex tasks produced an overall average of 1.40 for browsing search results.

Overall, reexamining search results (by the subjects) was fairly insignificant across these task categories, where combo-complex tasks produced a higher average for back browsing at only 0.55 per task. Once again, when comparing results based on the visual and textual classifications of the experimental search tasks, visual tasks demonstrated higher means. While easy task produced an overall mean of 0.19 for reexamining search results, easy-visual tasks scored a mean of 0.38 and easy-textual tasks actually averaged 0.00[103]. In addition, complex-visual tasks produced a mean score of 0.71, compared to complex-textual tasks at 0.19[104].

Subjects' use of the Details feature also indicated several differences across these task classifications. Combo-complex tasks produced higher levels of (Details) use at an overall mean of 2.69. Easy tasks, on the other hand, achieved a mean of 1.12 per task. Also, visual and textual task types again reflected differences among the use of the Details feature. For example, results for easy tasks revealed that easy-visual tasks produced a mean of 0.19, while easy-textual tasks showed a significantly greater average at 2.05.

Readers can also observe patterns for using the Promote features. Combo-complex tasks showed relatively high means for Promote searching at an overall average of 1.57. The average number of Promote searches for combination tasks closely followed at 1.40 per task. Complex tasks also produced a moderate average for Promote searching at 1.10. However, easy tasks showed a lower mean at 0.74. Again, the textual and visual

---

[103] An average score of 0.00 indicates that this technique for video browsing was never used across the entire experiment.

[104] These results combine for a composite score of 0.45 for complex tasks.

characteristics of easy search tasks[105] resulted in significant differences among scores, including a mean of 1.24 Promote searches for easy-visual tasks and 0.24 for easy-textual tasks.

During-test analysis also produced results for each of these interface features and functions across different system variants and task classifications. Readers can observe the composite means and standard deviations for each system in Table 20. To recap, System variant 1 excluded the Title Browse, System 2 excluded the keyword, or transcript, search, and System 3, otherwise referred to as the "full" system variant, implemented all search and browse features. (See Chapter 5 for review of all search and browse features implemented for the experimental version of ViewFinder.)

Composite results for System variant 1 revealed that subjects' use of the keyword search was, again, quite common at a rate of 2.75 searches per task (Table 20). The Details feature was also used rather frequently at 2.61 times per task. Other overall scores produced for System variant 1 included that browsing search results was performed an average of 1.45, while Promote searches were employed 1.07 times per task. Reexamining previously seen search results, i.e. browsing back through search results, once again exhibited low use at a mean of 0.27.

Results for System variant 2 differed from System 1 as the keyword (transcript) search was excluded, but the Title Browse was implemented. These results showed that the Title Browse feature was employed at a higher rates on System 2 than the other system variants at 2.66 times per task. The mean scores for browsing search results – while using System 2 – exhibited the next highest levels with an overall mean of 2.50.

[105] Easy tasks exhibited an overall mean of 0.74 Promote searches per task.

Results depicting the use of the Details feature showed a slight decrease from the averages produced by System 1 with an average score of 1.88. The number of Promote searches, however, did increase to 1.80 per task, and the averages for reexamining search results also raised to 0.68.

Experimental search runs employing System variant 3 were used to evaluate user interaction with the full system variant[106]. Results from System variant 3 exhibited some similarities with System 1. For one, keyword searches were performed an average of 2.39 times per task. In addition, analyzing subjects' tendency to browse and reexamine search results demonstrated somewhat similar findings when compared to the results for System variant 1. Moreover, the average number of times subjects browsed search results remained relatively high at 1.50 per task, while means for reexamining search results were low at 0.34. The use of the Title Browse feature significantly decreased from the results produced for System 2 and averaged 0.64. The number of Promote searches also slightly dropped when using System 3 to 0.73 searches per task.

Table 20 also presents results for how each individual system variant scored throughout different task types, such as textual, visual, and hybrid classifications of search tasks. According to this analysis, one key observation was that the use of the Details function on System 1 differed across these task categories. Variations among the use of the Details function were underscored by the averages produced for textual tasks at 4.71, compared to 0.79 for visual-only tasks. Keyword searching remained relatively constant across these different task categories; however, the number of Promote searches and reexamining search results did tend to vary. For example, Promote searches

[106] A "full system variant" implements all interface features and functions evaluated throughout this study.

averaged 0.57 on textual tasks, 1.36 on hybrid tasks, and 1.00 on visual-only tasks. Also, reexamining search results ranged from 0.00 for textual tasks to 0.50 for visual tasks.

By analyzing subjects' use of System variant 2, readers can observe further differences and associations among textual, visual, and hybrid tasks. For example, subjects' inclination to browse search results varied; System 2 demonstrated an overall mean of 0.93 on textual tasks compared to 3.86 for hybrid tasks. In addition, the number of Promote searches also differed across these task types. Textual search tasks produced a lower mean for Promote searching at 1.07 per task, while hybrid tasks scored an average of 2.29. Also, while using System 2, video (title) browsing remained relatively constant during hybrid tasks – at a mean of 3.18 per task – and visual tasks – at 2.14. The use of the Details feature did vary across these task categories where visual-only tasks averaged 2.57 and textual tasks only scored 0.64.

Results generated for System variant 3 showed additional findings. The Title Browse feature was used equally across textual, visual, and hybrid search tasks with a mean score of 0.64, apiece. In addition, Promote searching demonstrated approximately the same rate of use for visual and hybrid tasks at means of 0.86 and 0.82, respectively. Some distinctions among the use of System variant 3 derived from examining keyword searching and browsing search results. Moreover, subjects performed a mean of 3.07 keyword searches on visual tasks, compared to 1.43 on textual tasks[107]. When analyzing how subjects browsed search results on System 3, results again reflected greater usage for visual tasks with a mean of 2.86 per task. Textual tasks showed a mean 0.14 and hybrid tasks demonstrated a mean of 1.50 for browsing search results on System 3.

[107] Hybrid tasks scored in between visual and textual tasks at a mean of 2.54 keyword searches per task.

Table 20: Use of interface features and functions across different system variants and task types.

System	Task Type	Keyword Searches *(SD)*	Browse Video Titles (*SD)*	Browse More Search Results *(SD)*	Reexamine Search Results *(SD)*	View Clip Details *(SD)*	Promote Search Result *(SD)*
1	**Overall**	2.75 *(1.61)*	N/A	1.45 *(2.61)*	0.27 *(0.59)*	2.61 *(3.87)*	1.07 *(1.66)*
	Text-only	2.86 *(1.23)*	N/A	1.21 *(4.00)*	0.00 *(0.00)*	4.71 *(5.20)*	0.57 *(1.02)*
	Visual-only	2.43 *(1.51)*	N/A	1.64 *(2.10)*	0.50 *(0.76)*	0.79 *(1.31)*	1.00 *(1.18)*
	Hybrid	2.86 *(1.84)*	N/A	1.46 *(2.01)*	0.29 *(0.60)*	2.46 *(3.56)*	1.36 *(2.06)*
	Easy	2.29 *(1.27)*	N/A	2.07 *(3.97)*	0.29 *(0.61)*	2.71 *(5.47)*	1.14 *(1.29)*
	Easy-Textual	2.57 *(1.27)*	N/A	2.43 *(5.59)*	0.00 *(0.00)*	5.14 *(7.13)*	0.71 *(1.25)*
	Easy-Visual	2.00 *(1.29)*	N/A	1.71 *(1.60)*	0.57 *(0.79)*	0.29 *(0.49)*	1.57 *(1.27)*
	Complex	3.00 *(1.41)*	N/A	0.79 *(1.97)*	0.21 *(0.58)*	2.79 *(2.67)*	0.43 *(0.76)*
	Complex-Textual	3.14 *(1.22)*	N/A	0.00 *(0.00)*	0.00 *(0.00)*	4.29 *(2.69)*	0.43 *(0.79)*
	Complex-Visual	2.86 *(1.68)*	N/A	1.57 *(2.64)*	0.43 *(0.79)*	1.29 *(1.70)*	0.43 *(0.79)*
	Combination	1.86 *(1.03)*	N/A	1.14 *(2.35)*	0.21 *(0.58)*	1.00 *(2.04)*	1.50 *(2.68)*
	Combo-Complex	3.86 *(1.96)*	N/A	1.79 *(1.63)*	0.36 *(0.63)*	3.93 *(4.20)*	1.21 *(1.25)*
2	**Overall**	N/A	2.66 *(2.18)*	2.50 *(3.61)*	0.68 *(1.31)*	1.88 *(2.40)*	1.80 *(2.23)*
	Text-only	N/A	2.14 *(1.88)*	0.93 *(1.73)*	0.29 *(0.61)*	0.64 *(0.93)*	1.07 *(1.98)*
	Visual-only	N/A	2.14 *(2.07)*	1.36 *(1.87)*	0.36 *(0.93)*	2.57 *(3.78)*	1.57 *(1.95)*
	Hybrid	N/A	3.18 *(2.31)*	3.86 *(4.43)*	1.04 *(1.62)*	2.14 *(2.22)*	2.29 *(2.42)*
	Easy	N/A	1.21 *(0.89)*	0.36 *(0.93)*	0.00 *(0.00)*	0.07 *(0.27)*	0.71 *(1.14)*
	Easy-Textual	N/A	1.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*
	Easy-Visual	N/A	1.43 *(1.27)*	0.71 *(1.25)*	0.00 *(0.00)*	0.14 *(0.38)*	1.43 *(1.27)*
	Complex	N/A	3.07 *(2.27)*	1.93 *(2.09)*	0.64 *(1.01)*	3.14 *(2.91)*	1.93 *(2.40)*
	Complex-Textual	N/A	3.29 *(2.14)*	1.86 *(2.12)*	0.57 *(0.79)*	1.29 *(0.95)*	2.14 *(2.41)*
	Complex-Visual	N/A	2.86 *(2.55)*	2.00 *(2.24)*	0.71 *(1.25)*	5.00 *(3.06)*	1.71 *(2.56)*
	Combination	N/A	3.21 *(2.99)*	3.07 *(3.95)*	1.14 *(2.03)*	2.29 *(2.56)*	2.07 *(1.73)*
	Combo-Complex	N/A	3.14 *(1.46)*	4.64 *(4.88)*	0.93 *(1.14)*	2.00 *(1.92)*	2.50 *(3.00)*
3	**Overall**	2.39 *(1.53)*	0.64 *(0.88)*	1.50 *(1.99)*	0.34 *(0.79)*	1.75 *(2.05)*	0.73 *(1.15)*
	Text-only	1.43 *(1.28)*	0.64 *(0.50)*	0.14 *(0.54)*	0.00 *(0.00)*	1.36 *(1.60)*	0.43 *(0.85)*
	Visual-only	3.07 *(1.82)*	0.64 *(0.63)*	2.86 *(1.70)*	0.79 *(1.12)*	1.21 *(1.85)*	0.86 *(1.35)*
	Hybrid	2.54 *(1.29)*	0.64 *(1.13)*	1.50 *(2.15)*	0.29 *(0.71)*	2.21 *(2.28)*	0.82 *(1.19)*
	Easy	1.64 *(1.95)*	0.71 *(0.61)*	1.50 *(1.82)*	0.29 *(0.61)*	0.57 *(1.28)*	0.36 *(0.63)*
	Easy-Textual	0.57 *(1.13)*	0.71 *(0.49)*	0.00 *(0.00))*	0.00 *(0.00)*	1.00 *(1.73)*	0.00 *(0.00)*
	Easy-Visual	2.71 *(2.06)*	0.71 *(0.76)*	3.00 *(1.41)*	0.57 *(0.79)*	0.14 *(0.38)*	0.71 *(0.76)*
	Complex	2.86 *(1.35)*	0.57 *(0.51)*	1.50 *(1.95)*	0.50 *(1.09)*	2.00 *(1.80)*	0.93 *(1.44)*
	Complex-Textual	2.29 *(0.76)*	0.57 *(0.54)*	0.29 *(0.76)*	0.00 *(0.00)*	1.71 *(1.50)*	0.86 *(1.07)*
	Complex-Visual	3.43 *(1.62)*	0.57 *(0.54)*	2.71 *(2.06)*	1.00 *(1.41)*	2.29 *(2.14)*	1.00 *(1.83)*
	Combination	2.07 *(1.27)*	1.00 *(1.47)*	1.36 *(1.55)*	0.21 *(0.58)*	2.29 *(2.59)*	0.64 *(0.75)*
	Combo-Complex	3.00 *(1.18)*	0.29 *(0.47)*	1.64 *(2.68)*	0.36 *(0.84)*	2.14 *(2.03)*	1.00 *(1.52)*

Readers can also observe (Table 20) how subjects interacted with the different system variants across other task classifications designed for this study[108]. One interesting observation included that System variant 1 produced a higher frequency for keyword searching when used for combo-complex tasks at an average of 3.86. These results can be compared to the results for combination tasks, on System 1, which scored a mean of 1.86 keyword searches per task. In regards to browsing search results, easy tasks exhibited a higher rate at 2.07 per task, while complex tasks showed a lower mean of 0.79. It's also important to examine differences in browsing search results during complex tasks specifically, which scored an overall mean of 0.79. Results for complex tasks showed that complex-visual tasks averaged 1.57 for browsing search results, compared to complex-textual tasks at a mean of 0.00. By examining the use of the Promote search feature, combination tasks demonstrated higher levels with a mean score of 1.50 per task. Results for complex tasks reflected a lower average for Promote searching at 0.43. The use of the Details feature – as evaluated for System 1 – also exhibited relatively high averages across multiple task categories. Means scores for using the Details feature ranged from 3.93 for combo-complex tasks to 2.71 for easy tasks. Once again, it is important to examine differences among easy and complex tasks, according to the visual and textual characteristics of each task type. Easy-textual tasks demonstrated significantly higher use of the Details feature than easy-visual tasks with mean scores of 5.14 and 0.29, respectively. A similar trend was also observed when analyzing complex-textual and complex-visual tasks; complex-textual tasks demonstrated

[108] The other task categories include easy, complex, combination, and combo-complex tasks.

higher use of the Details feature than complex-visual tasks with a mean of 4.29 compared to 1.29.

Next, readers can examine differences among these task categories and interaction with System variant 2. First, results of the Title Browse feature were shown to be somewhat constant across several of these task categories, including combination tasks scoring 3.21, combo-complex tasks at 3.14, and complex tasks at 3.07. There were other interesting observations when analyzing the use of the Promote features on System 2. Moreover, combo-complex tasks demonstrated a higher average for Promote searching at 2.50 times per task, and combination and complex tasks closely followed at 2.07 and 1.93, respectively. On the other hand, while easy tasks averaged 0.71 Promote searches per task, easy-visual tasks actually exhibited a mean of 1.43 and easy-textual tasks scored 0.00. Results for System variant 2 demonstrated other interesting patterns among the use of the Details feature. Overall, complex tasks exhibited higher use of the Details feature at a mean of 3.14; however, complex-visual tasks averaged 5.00 and complex-textual tasks scored 1.29. Other results depicting the use of the Details feature included that combination tasks averaged 2.29 and easy tasks produced an extremely low mean of 0.07. Other interesting findings (or interactions) for System variant 2 were also observed from examining subjects' inclination to browse search results. For example, combo-complex tasks scored a high mean for browsing search results at 4.64, compared to easy tasks at a low of 0.36. Easy-textual tasks averaged 0.00 for browsing search results; however, easy-visual tasks did present minimal levels at 0.71. Other scores produced for browsing search results included that combination and complex tasks produced moderate to high scores, on average, at 3.07 and 1.93, respectively.

Table 20 also presents an analysis of the interface features and functions implemented for System variant 3. By examining these results, keyword searching was shown to be somewhat constant across various task categories, including means that ranged from 3.00 for combo-complex tasks to 1.64 for easy tasks. Other scores depicting keyword searching on System 3 were also moderately high, including complex tasks at an overall mean of 2.86. Complex-visual tasks exhibited more keyword searches than complex-textual at a rate of 3.43 per task, compared to 2.29. Also, easy-textual tasks exhibited significantly less keyword searches when compared to easy-visual tasks[109] with a mean of only 0.57.

Other trends can be observed among the use of browse features on System variant 3. Scores for the Title Browse were relatively low across these task categories, including means that varied from 1.00 for combination tasks to 0.29 for combo-complex tasks. Other interesting trends were apparent when examining subjects' tendency to browse search results on System 3. While easy and complex tasks produced the same at overall score of 1.50, again, the influence of visual and textual attributes of these task categories were evident. Moreover, easy-visual and complex-visual tasks averaged significantly higher levels of (search) results browsing than easy-textual and complex-textual tasks with mean scores of 3.00 and 2.71, compared to 0.00 and 0.29, respectively. This pattern can also be observed when analyzing how subjects reexamined search results; easy-visual and complex-visual tasks demonstrated higher averages for reexamining search results than easy-textual and complex-textual search tasks.

[109] A mean of 2.71 keyword searches per task was observed.

In addition, analysis of System variant 3 produced fairly insignificant results for Promote searching. Combo-complex tasks demonstrated a rate for Promote searching at 1.00, while scores for complex, combination, and easy tasks ranged from 0.93 to 0.36. Easy-textual tasks actually averaged 0.00 Promote searches per task – while using System 3 – but easy-visual tasks did achieve a mean of 0.71. These scores signified differences in user behavior across visual and textual types of search tasks.

Analysis of during-test results also allowed a closer examination of subjects' use of individual Promote search functions. This analysis was important because results depicted how subjects interacted with visual search features and another variation of textual searching, or the Textual Promote. To recap, the visual search features analyzed as part of this study included promoting search results, or keyframes, by color, shape, texture, and all-visual[110] attributes. Results for the Promote search features are presented in Tables 21 and 22. These results have been analyzed in a way that allows readers to examine how subjects used the visual and textual Promote features overall and across different task types and system variants.

By examining the results in Table 21, readers can observe the overall use of each Promote search feature. Composite results – or the collective means of all search runs – can be found in the row marked *Overall*. These results presented that, across all system variants and task types, the Textual Promote was used 0.46 times per task while a visual search was performed an average of 0.65 times. These composite results also show the average use of all individual visual searches. This analysis demonstrated that color was

[110] The All-Visuals Promote search used a combination of color, shape, and texture attributes to retrieve results.

the visual attribute most frequently promoted, at a mean of 0.38 per task. Other scores demonstrated that the All-Visuals Promote was performed the second highest, at 0.15, followed by the Hybrid Promote at 0.11. The Shape and Texture Promote features exhibited little influence throughout the full experiment as each only demonstrated an average use of 0.08 and 0.04, per task, respectively.

Table 21: Use of promote searches overall and among different task types.

Task Types	Textual Promote (*SD*)	Visual Promote (*SD*)	Hybrid Promote (*SD*)	Color Promote (*SD*)	Shape Promote (*SD*)	Texture Promote (*SD*)	All-Visuals Promote (*SD*)
Overall	0.46 *(0.99)*	0.65 *(1.17)*	0.11 *(0.40)*	0.38 *(0.82)*	0.08 *(0.40)*	0.04 *(0.22)*	0.15 *(0.44)*
Text-only	0.52 *(0.94)*	0.14 *(0.52)*	0.02 *(0.15)*	0.02 *(0.15)*	0.07 *(0.34)*	0.00 *(0.00)*	0.05 *(0.22)*
Visual-only	0.21 *(0.52)*	0.81 *(1.19)*	0.12 *(0.40)*	0.52 *(0.86)*	0.07 *(0.26)*	0.05 *(0.22)*	0.17 *(0.38)*
Hybrid	0.55 *(1.17)*	0.83 *(1.33)*	0.14 *(0.47)*	0.49 *(0.94)*	0.10 *(0.48)*	0.05 *(0.27)*	0.20 *(0.53)*
Easy	0.17 *(0.49)*	0.50 *(0.80)*	0.07 *(0.26)*	0.38 *(0.66)*	0.05 *(0.22)*	0.00 *(0.00)*	0.07 *(0.26)*
Easy-Textual	0.14 *(0.48)*	0.05 *(0.22)*	0.05 *(0.22)*	0.00 *(0.00)*	0.05 *(0.22)*	0.00 *(0.00)*	0.00 *(0.00)*
Easy-Visual	0.19 *(0.51)*	0.95 *(0.92)*	0.10 *(0.30)*	0.76 *(0.77)*	0.05 *(0.22)*	0.00 *(0.00)*	0.14 *(0.36)*
Complex	0.57 *(0.94)*	0.45 *(1.13)*	0.07 *(0.34)*	0.17 *(0.66)*	0.10 *(0.37)*	0.05 *(0.22)*	0.14 *(0.35)*
Complex-Textual	0.90 *(1.14)*	0.24 *(0.70)*	0.00 *(0.00)*	0.05 *(0.22)*	0.10 *(0.44)*	0.00 *(0.00)*	0.10 *(0.30)*
Complex-Visual	0.24 *(0.54)*	0.67 *(1.43)*	0.14 *(0.48)*	0.29 *(0.90)*	0.10 *(0.30)*	0.10 *(0.30)*	0.19 *(0.40)*
Combination	0.52 *(1.02)*	0.79 *(1.30)*	0.10 *(0.30)*	0.38 *(0.99)*	0.07 *(0.26)*	0.05 *(0.31)*	0.29 *(0.64)*
Combo-Complex	0.57 *(1.31)*	0.88 *(1.37)*	0.19 *(0.59)*	0.60 *(0.89)*	0.12 *(0.63)*	0.05 *(0.22)*	0.12 *(0.40)*

During-test results were also evaluated to explore the use of individual Promote features across textual, visual, and hybrid search tasks. When assessing the Textual Promote feature, hybrid and text-only tasks demonstrated higher use at 0.55 and 0.52 per task. On the other hand, textual (Promote) searching was not performed as often on visual-only tasks, with a mean score of 0.21. Visual searching varied, as results indicated higher use overall among hybrid (M=0.83) and visual tasks (M=0.81). On the other hand, textual tasks exhibited significantly lower levels of visual searching at only 0.14 per task. When closely examining the use of the Color Promote on textual, visual, and hybrid tasks, other interesting observations were made. Results showed that color searching was performed, on average, a mean of 0.52 for visual tasks, followed by 0.49 for hybrids tasks and only 0.02 for textual-only tasks. Hybrid, Texture, and Shape Promote features, again, produced insignificant results across these task types.

Table 21 presents the use of Promote features across additional classifications of search tasks, such as easy, complex, combination, and combo-complex tasks. Interaction with the Textual Promote indicated that complex and combo-complex tasks resulted in relatively higher scores at 0.57. By further examining the results for complex tasks, readers can observe how complex-textual tasks resulted in greater use of the Textual Promote than complex-visual tasks, with a mean score of 0.90 compared to 0.24. Easy tasks demonstrated lower averages for the Textual Promote feature and also didn't present any clear distinctions between the results for visual and textual forms of (easy) tasks.

Readers can also examine subjects' interaction with the visual Promote features throughout these task types. First, results showed that visual searching was performed more often on combo-complex tasks, at an average of 0.88 per task, and combination tasks, at a mean of 0.79, than during the other task categories. One important observation from this analysis was that easy tasks averaged 0.50 visual searches per task; however, easy-visual tasks actually produced a mean of 0.95 and easy-textual tasks exhibited a significantly lower average at 0.05. Visual searches among these task categories also indicated that color searching was performed more frequently on combo-complex tasks (M=0.60) and easy-visual tasks (M=0.76). These results contrast with easy-textual tasks, which averaged 0.00 color searches per task. Once again, results across easy, complex, combination, and combo-complex tasks demonstrated that the Shape, Texture, and Hybrid Promote features were fairly insignificant.

Table 22: Use of promote searches across different system variants and task types.

System	Task Types	Textual Promote (*SD*)	Visual Promote (*SD*)	Hybrid Promote (*SD*)	Color Promote (*SD*)	Shape Promote (*SD*)	Texture Promote (*SD*)	All-Visuals Promote (*SD*)
1	**Overall**	0.41 *(0.89)*	0.57 *(1.01)*	0.09 *(0.35)*	0.27 *(0.52)*	0.05 *(0.23)*	0.05 *(0.30)*	0.20 *(0.59)*
	Text-only	0.43 *(0.76)*	0.07 *(0.27)*	0.07 *(0.67)*	0.00 *(0.00)*	0.07 *(0.27)*	0.00 *(0.00)*	0.00 *(0.00)*
	Visual-only	0.21 *(0.43)*	0.71 *(0.91)*	0.07 *(0.27)*	0.43 *(0.65)*	0.07 *(0.27)*	0.07 *(0.27)*	0.14 *(0.36)*
	Hybrid	0.50 *(1.11)*	0.75 *(1.21)*	0.11 *(0.45)*	0.32 *(0.55)*	0.04 *(0.19)*	0.07 *(0.38)*	0.32 *(0.77)*
	Easy	0.36 *(0.63)*	0.64 *(0.93)*	0.14 *(0.36)*	0.43 *(0.65)*	0.14 *(0.36)*	0.00 *(0.00)*	0.07 *(0.27)*
	Easy-Textual	0.43 *(0.79)*	0.14 *(0.38)*	0.14 *(0.38)*	0.00 *(0.00)*	0.14 *(0.38)*	0.00 *(0.00)*	0.00 *(0.00)*
	Easy-Visual	0.29 *(0.48)*	1.14 *(1.07)*	0.14 *(0.38)*	0.86 *(0.69)*	0.14 *(0.38)*	0.00 *(0.00)*	0.14 *(0.38)*
	Complex	0.29 *(0.61)*	0.14 *(0.36)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.07 *(0.27)*	0.07 *(0.27)*
	Complex-Textual	0.43 *(0.79)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*
	Complex-Visual	0.14 *(0.38)*	0.29 *(0.49)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.14 *(0.38)*	0.14 *(0.38)*
	Combination	0.57 *(1.40)*	0.86 *(1.46)*	0.07 *(0.27)*	0.21 *(0.43)*	0.07 *(0.27)*	0.14 *(0.54)*	0.43 *(0.94)*
	Combo-Complex	0.43 *(0.76)*	0.64 *(0.93)*	0.14 *(0.54)*	0.43 *(0.65)*	0.00 *(0.00)*	0.00 *(0.00)*	0.21 *(0.58)*
2	**Overall**	0.73 *(1.30)*	0.98 *(1.60)*	0.09 *(0.29)*	0.61 *(1.16)*	0.16 *(0.63)*	0.04 *(0.19)*	0.18 *(0.39)*
	Text-only	0.79 *(1.31)*	0.29 *(0.83)*	0.00 *(0.00)*	0.07 *(0.27)*	0.14 *(0.54)*	0.00 *(0.00)*	0.07 *(0.27)*
	Visual-only	0.43 *(0.76)*	1.07 *(1.69)*	0.07 *(0.27)*	0.64 *(1.08)*	0.07 *(0.27)*	0.07 *(0.27)*	0.29 *(0.47)*
	Hybrid	0.86 *(1.51)*	1.29 *(1.78)*	0.14 *(0.36)*	0.86 *(1.38)*	0.21 *(0.79)*	0.04 *(0.19)*	0.18 *(0.39)*
	Easy	0.14 *(0.54)*	0.57 *(0.94)*	0.00 *(0.00)*	0.43 *(0.85)*	0.00 *(0.00)*	0.00 *(0.00)*	0.14 *(.363)*
	Easy-Textual	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*
	Easy-Visual	0.29 *(0.76)*	1.14 *(1.07)*	0.00 *(0.00)*	0.86 *(1.07)*	0.00 *(0.00)*	0.00 *(0.00)*	0.29 *(0.49)*
	Complex	1.07 *(1.27)*	0.79 *(1.72)*	0.07 *(0.27)*	0.29 *(0.83)*	0.21 *(0.58)*	0.07 *(0.27)*	0.21 *(0.43)*
	Complex-Textual	1.57 *(1.51)*	0.57 *(1.13)*	0.00 *(0.00)*	0.14 *(0.38)*	0.29 *(0.76)*	0.00 *(0.00)*	0.14 *(0.38)*
	Complex-Visual	0.57 *(0.79)*	1.00 *(2.24)*	0.14 *(0.38)*	0.43 *(1.13)*	0.14 *(0.38)*	0.14 *(0.38)*	0.29 *(0.48)*
	Combination	0.79 *(0.89)*	1.07 *(1.64)*	0.21 *(0.43)*	0.71 *(1.59)*	0.14 *(0.36)*	0.00 *(0.00)*	0.21 *(0.43)*
	Combo-Complex	0.93 *(1.98)*	1.50 *(1.95)*	0.07 *(0.27)*	1.00 *(1.18)*	0.29 *(1.07)*	0.07 *(0.27)*	0.14 *(0.36)*
3	**Overall**	0.23 *(0.60)*	0.41 *(0.65)*	0.14 *(0.52)*	0.27 *(0.59)*	0.04 *(0.19)*	0.02 *(0.13)*	0.09 *(0.29)*
	Text-only	0.36 *(0.63)*	0.07 *(0.27)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.07 *(0.27)*
	Visual-only	0.00 *(0.00)*	0.64 *(0.84)*	0.21 *(0.58)*	0.50 *(0.86)*	0.07 *(0.27)*	0.00 *(0.00)*	0.07 *(0.27)*
	Hybrid	0.29 *(0.71)*	0.46 *(0.64)*	0.18 *(0.61)*	0.29 *(0.54)*	0.04 *(0.19)*	0.04 *(0.19)*	0.11 *(0.32)*
	Easy	0.00 *(0.00)*	0.29 *(0.47)*	0.07 *(0.27)*	0.29 *(0.47)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*
	Easy-Textual	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*
	Easy-Visual	0.00 *(0.00)*	0.57 *(0.54)*	0.14 *(0.38)*	0.57 *(0.54)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*
	Complex	0.36 *(0.63)*	0.43 *(0.85)*	0.14 *(0.54)*	0.21 *(0.80)*	0.07 *(0.27)*	0.00 *(0.00)*	0.14 *(0.36)*
	Complex-Textual	0.71 *(0.76)*	0.14 *(0.38)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.00 *(0.00)*	0.14 *(0.38)*
	Complex-Visual	0.00 *(0.00)*	0.71 *(1.11)*	0.29 *(0.76)*	0.43 *(1.13)*	0.14 *(0.38)*	0.00 *(0.00)*	0.14 *(0.38)*
	Combination	0.21 *(0.58)*	0.43 *(0.51)*	0.00 *(0.00)*	0.21 *(0.43)*	0.00 *(0.00)*	0.00 *(0.00)*	0.21 *(0.43)*
	Combo-Complex	0.36 *(0.84)*	0.50 *(0.76)*	0.36 *(0.84)*	0.36 *(0.63)*	0.07 *(0.27)*	0.07 *(0.27)*	0.00 *(0.00)*

The final reporting of during-test results included evaluating the use of individual Promote search functions across different task categories and system variants. Means and standard deviations from this analysis are reported in Table 22.

By examining the results produced for each system variant, readers can observe the levels of textual, visual, and hybrid (Promote) searching throughout the experimental search runs. First, Table 22 presents during-test results for System variant 1. Results for System 1 demonstrated that visual searching was employed more often than the Textual Promote with an overall mean of 0.57, compared to 0.41. In addition, subjects' use of System 1 indicated that the Color Promote was one of the more commonly performed visual searches at an overall average of 0.27 per task. In addition, the All-Visuals Promote was used an average of 0.20 on System variant 1.

Results depicting the use of the Promote features on System variant 2, again, indicated that visual searching was performed more often than textual (Promote) searching. Moreover, the number of visual searches averaged 0.98 per task while the Textual Promote was used a mean of 0.73 times. Results produced for System variant 2 also demonstrated that the Color Promote was used more frequently than any other visual search feature with a mean of 0.61[111].

Composite results for System 3, again, revealed that visual searches were performed more often than a Textual Promote, with mean scores of 0.41 and 0.23, respectively. In addition, the Color Promote was shown to be used more often than any other visual search feature at a mean of 0.27 times per tasks. The next highest scores for

[111] Results for the Color Promote on System 2 can be compared to the means for Shape Promote at 0.16, Texture Promote at 0.04, and the All-Visuals at 0.18.

a visual search feature, when evaluating System 3, included the Hybrid Promote, which produced a mean of 0.14. Subjects' use of the Shape and Texture Promote features did not produce any significant results for searching behaviors.

This study also analyzed individual Promote features and their relation to different system variants and task types, collectively. First, this analysis allows readers to examine differences among the use of various Promote features for textual, visual, and hybrid search tasks while using System variant 1. As shown in Table 22, Textual Promote searches were performed somewhat similarly across these different task categories. Results for System 1 presented that hybrid tasks averaged 0.50, textual tasks scored 0.43, and visual tasks measured 0.21 (Textual Promote) searches per task.

Results for visual searching on System 1 were more distinct. Visual searches were performed, on average, 0.75 times for hybrid tasks and 0.71 for visual tasks. The number of visual searches on text-only tasks was significantly different at only 0.07 per task. The rates, of which certain visual features were employed across these task classifications, while using System 1, presented other findings. The Color Promote feature was used more often for visual tasks (M=0.43) and hybrid search tasks (M=0.32). On the other hand, 0.00 color, texture, and all-visual searches were performed for text-only tasks.

Readers can also examine results for System variant 2. First, there were minimal differences in how the Textual Promote feature was employed across each of these task categories. Hybrid tasks showed a mean of 0.86 (Textual Promote) searches per task, while textual tasks scored 0.79 and visual tasks averaged 0.43.

Results revealed greater distinctions for visual searching across textual, visual, and hybrid search tasks. Visual searches were performed an average of 1.29 times for hybrid tasks and 1.07 times for visual tasks, scores that indicated recurrent use. However, textual tasks only produced an average of 0.29 visual searches per task. Other notable observations were, again, made in regards to color searching. Moreover, hybrid tasks averaged 0.86 color searches per task – while using System variant 2 – and visual tasks produced a mean of 0.64. Text-only tasks, on the other hand, only achieved 0.07 color searches per task. Less-significant results were observed after evaluating the Hybrid, Shape, Texture, and All-Visual Promote features for System 2. Text-only tasks actually demonstrated a mean of 0.00 hybrid and texture searches. However, the All-Visuals feature was used an average of 0.29 times for visual-only tasks, and shape searches averaged 0.21 on hybrid tasks.

Use of the Promote features on System variant 3 was also analyzed for textual, visual, and hybrid tasks. Results from the Textual Promote differed across these task types. For example, text-only and hybrid search tasks produced higher rates of Textual Promote searching at means of 0.36 and 0.29, respectively. Visual-only tasks, on the other hand, actually resulted in 0.00 Textual Promote searches over all search runs.

Patterns for visual searching also varied across some of these task categories. Visual-only tasks achieved a mean of 0.64 visual searches per tasks, while hybrid tasks produced a mean of 0.46, and textual tasks averaged the lower rate for visual searching at 0.07. One of the more distinct observations among Promote searching on System 3 was that several visual features achieved similar scores during text-only search tasks. Moreover, every Visual Promote feature, with the exception of the All-Visuals Promote

(M=0.07), scored exactly 0.00 for text-only search tasks. On the other hand, visual-only tasks achieved a moderate average for the Color Promote, at a mean of 0.50, and Hybrid Promote, at 0.21. Hybrid tasks exhibited that color searches were performed 0.29 times per task, and hybrid searches were employed 0.18 times.

Similar to previous analyses of during-test results, readers can explore distinctions among Promote features and how they related to easy, complex, combination, and combo-complex search tasks and different system variants. This analysis (Table 22) first depicted the use of various Promote features implemented for System 1. The Textual Promote feature (on System 1) demonstrated higher averages for combination tasks, at a mean of 0.57. The scores for combination tasks precede those for combo-complex tasks, which averaged 0.43 textual (Promote) searches per task. Complex tasks produced an overall mean for the Textual Promote feature at 0.29, but also revealed some interesting trends among the textual and visual characteristics of search tasks and their influence on searching behaviors. Moreover, there were distinctions between subjects' use of the Textual Promote on complex-textual and complex-visual search tasks; complex-textual tasks exhibited 0.43 textual (Promote) searches per task while complex-visual averaged 0.14.

Differences among visual searching, while using System 1, were also revealed from this analysis. Overall, Visual Promote searches were employed somewhat regularly across these task categories. Visual searching was performed more frequently on combination tasks, at a mean of 0.86 per task, followed by easy and combo-complex tasks, at 0.64 apiece. Results from the easy tasks presented clear distinctions between easy-textual and easy-visual tasks, where easy-visual tasks averaged 1.14 visual searches per task, compared to 0.14 for easy-textual tasks.

Other interesting trends were observed from examining the use of specific visual search features on System variant 1. For example, the Hybrid, Color, and Shape Promote searches were employed a total of 0.00 times across all complex search tasks. In addition, the Texture Promote feature was never used during any easy task, including easy-textual or easy-visual tasks. Complex-textual tasks also totaled 0.00 visual searches, of any kind, while using System 1. Other, more positive, results for visual searching on System 1 included a mean of 0.43 color searches for both easy and combo-complex search tasks. When closely examining easy search tasks, and their relation to the use of the Color Promote on System 1, easy-visual tasks averaged a mean of 0.86 while easy-textual tasks scored 0.00[112]. Easy-textual tasks also produced a mean of 0.00 for Texture and All-Visuals (Promote) searching on System 1.

Next, readers can explore how individual Promote features were used while searching these various task categories on System variant 2. First, let's examine the use of the Textual Promote on System 2. When analyzing these specific task categories, complex tasks averaged higher means for textual (Promote) searching at 1.07; these scores contrast with the results for easy tasks, which exhibited 0.14 per task. Combo-complex and combination tasks also achieved relatively positive means for the Textual Promote, and each task averaged 0.93 and 0.79, respectively. Upon closer examination,

[112] One possible explanation for such a result could be a certain degree of task bias. Future analyses will address and justify implications for potential task biases within these results and additional rounds of search experiments.

complex tasks highlighted other differences among the use of System 2, as results presented that complex-textual tasks produced a significantly higher mean for the Textual Promote feature at 1.57, compared to complex-visual at 0.57.

Subjects' tendency to perform visual searches during these task types, on System variant 2, was also analyzed. Combo-complex tasks exhibited recurrent use of the Visual Promote features at a mean 1.50 per task, while combination tasks closely followed at a rate of 1.07. Easy tasks, which achieved an overall mean of 0.57 visual (Promote) searches on System 2, again, demonstrated differences between textual and visual search tasks and their influence on searching behaviors. Moreover, easy-textual tasks actually averaged 0.00 visual searches per task and easy-visual tasks achieved a fairly high mean for visual searching at 1.14.

Analyzing specific visual queries, employed on System 2, revealed other findings. The Hybrid, Shape, and Texture Promote features were never used on any easy task. In addition, easy-textual tasks averaged 0.00 Color and All-Visuals Promote searches as well. On the other hand, while there were no color searches performed for any easy-textual tasks, easy-visual task averaged 0.86 per task. Combo-complex tasks scored an even higher mean for color searching at 1.00, which is also slightly higher than combination tasks, at mean of 0.71 per task.

Lastly, during-test results were evaluated to depict the use of individual Promote features on System variant 3. Readers can first observe how the Textual Promote was employed during easy, complex, combination, and combo-complex search tasks. Results for System 3 showed that textual (Promote) searching was never performed on any easy task, including any easy-textual or easy-visual tasks. These results also presented that complex and combo-complex tasks both produced an average of 0.36 Textual Promote searches per task. By closely examining complex tasks, in relation to System variant 3, it becomes apparent that complex-textual tasks averaged 0.71 text promotes per task, while complex-visual tasks actually produced a mean of 0.00.

Visual searching on System variant 3 was also analyzed. Subjects' interaction with the visual search features on System 3 showed higher frequencies during combo-complex tasks, at a mean of 0.50, and combination and complex task, at 0.43, apiece. While complex tasks demonstrated an average of 0.43 and easy tasks produced a mean a 0.29, results of these two task types presented clear distinctions between visual searching on visual and textual forms of search tasks. Moreover, the results for System 3 exhibited that complex-visual and easy-visual tasks averaged 0.71 and 0.57 visual searches, respectively, while the textual forms of these two task categories scored significantly lower at 0.00 and 0.14.

Evaluation of the individual Visual Promote features revealed other findings. Moreover, no (M=0.00) Shape, Texture, or All-Visuals Promote searches were ever performed on any easy task. The Hybrid and Color Promote features demonstrated a mean of 0.00 for easy-textual tasks, as well. Other visual search features, which also averaged 0.00 on System 3 included: the Texture Promote during both complex and combination tasks, the Shape and Hybrid Promote on complex-textual and combination tasks, and the All-Visuals Promote on combo-complex tasks. Readers can also observe the use of the Color Promote on System 3 and the differences among the visual and textual forms of easy and complex task types. For example, while easy and complex tasks averaged 0.29 and 0.21 color searches per task, respectively, each of the textual forms these task categories, i.e. easy-textual and complex-textual tasks, scored means of 0.00. On the other hand, complex-visual and easy-visual tasks scored significantly higher levels for color searching at 0.43 and 0.57, respectively.

6.3 Subjective Results

This study also incorporated subjective analysis. Subjective results were used to evaluate subjects' responses on the post-search (Appendix O) and post-experiment questionnaires (Appendix P), and short interviews concluding each search experiment (Appendix Q). The subjective analysis of this study assessed various experimental factors related to search tasks, system effectiveness and usability. Subjective results collected via the post-search and post-experiment questionnaires were based on a 5-point scale where "5" indicated highly favorable and "1" indicated unfavorable. Results from the subjective analysis are summarized in Tables 23 through 37.

6.3.1 Post-Search Results

Tables 23 through 26 present subjects' responses, i.e. average scores, deriving from the post-search questionnaires. Tables 23 and 24, specifically, report the means for certain factors related to the experimental search tasks (Appendix M). Tables 25 and 26 summarize subjects' assessment of system effectiveness, including interface functionality, system usability, and retrieval.

Composite results of the task-related factors, as collected on the post-search questionnaire, are presented in Table 23. First, subjects were asked to judge their familiarity with the topical nature of each search task. In addition, subjects rated how representative each task was of "real" video search tasks encountered by K – 12 science educators. Next, subjects estimated the efforts required[113] for searching each task and also assessed their individual performance for task completion. Closer examination of these results (Table 23) allows readers to observe the means and standard deviations for each of these tasks-related factors, both overall and across specific task types.

Table 23: Mean scores for post-search questions relating to search tasks.

Task Type	Familiarity with Search Task, Topic *(SD)*	Representation of Search Task (*SD*)	Ease of Searching on Topic *(SD)*	User Assessment of Task Completion *(SD)*
Overall	2.82 *(1.26)*	3.59 *(1.13)*	3.45 *(1.37)*	3.48 *(1.54)*
Text-only	2.40 *(1.25)*	3.43 *(1.25)*	3.74 *(1.45)*	3.67 *(1.53)*
Visual-only	3.50 *(1.24)*	3.71 *(1.15)*	3.81 *(1.13)*	3.86 *(1.28)*
Hybrid	2.68 *(1.14)*	3.61 *(1.05)*	3.13 *(1.37)*	3.20 *(1.63)*
Easy	2.98 *(1.39)*	3.48 *(1.23)*	4.14 *(1.05)*	4.07 *(1.05)*
Easy Textual	2.33 *(1.11)*	3.57 *(1.25)*	4.14 *(1.24)*	4.00 *(1.27)*
Easy Visual	3.62 *(1.36)*	3.38 *(1.24)*	4.14 *(0.85)*	4.14 *(0.85)*
Complex	2.93 *(1.33)*	3.67 *(1.18)*	3.40 *(1.42)*	3.45 *(1.42)*
Complex Textual	2.48 *(1.40)*	3.29 *(1.27)*	3.33 *(1.56)*	3.33 *(1.71)*
Complex Visual	3.38 *(1.12)*	4.05 *(0.97)*	3.48 *(1.29)*	3.57 *(1.57)*
Combination	2.69 *(1.18)*	3.93 *(0.89)*	3.17 *(1.32)*	3.29 *(1.61)*
Combo-Complex	2.67 *(1.12)*	3.29 *(1.11)*	3.10 *(1.43)*	3.12 *(1.66)*

The scores computed for all task types and system variants can be found in the row marked *Overall* (Table 23). These *Overall* scores reflected moderate assessment, or favorableness, for each of the task-related factors. For example, subjects felt the search tasks were reasonably reflective of actual search tasks found throughout K – 12 Science

[113] The effort required for each task was collected by asking subjects to rate, "ease of searching."

Education, at an average of 3.59 out of 5.0. In addition, moderate to high averages were produced for the ease of searching and completion of each search task, with mean scores of 3.45 and 3.48, respectively. Subjects were also shown to be somewhat familiar with the topic of each search task at a mean of 2.82 out of 5.0.

Table 23 also presents mean scores for these subjective (task-related) factors, as each scored across textual, visual, and hybrid task types. From these results, readers can first observe differences among subjects' familiarity with the search tasks. Subjects exhibited higher familiarity for visual-only tasks, at a mean of 3.50, compared to hybrid and text-only tasks, which averaged of 2.68 and 2.40, respectively. Results depicting subjects' assessment of task representation were more constant in that visual tasks were shown to be more reflective of actual K – 12 Science Education tasks, at a mean score of 3.71. Subjects also felt that textual tasks were only slightly less reflective of "real" tasks (M=3.43). Additional, more-constant, scores were also observed when measuring the ease of searching each search task. These task types, i.e. textual, visual, and hybrid tasks, were all shown to be relatively straightforward with visual tasks achieving a mean of 3.81, text-only tasks averaging 3.74, and hybrid tasks scoring 3.13 out of 5.0. Subjects' self-assessment of task completion also demonstrated somewhat constant scores where visual-only tasks exhibited a higher average at 3.86, with textual tasks at 3.67 and hybrid tasks at 3.20.

Readers can also examine (Table 23) subjects' responses on these post-search (task-related) questions as they were scored during easy, complex, combination, and combo-complex search tasks. First, subjects exhibited higher familiarity with easy and complex tasks at averages of 2.98 and 2.93, respectively. By closely examining results for easy and complex tasks, results also revealed distinctions between the visual and textual forms of these task types. Subjects' familiarity with easy and complex search tasks tended to increase for visual tasks, including mean scores of 3.62 for easy-visual tasks and 3.38 for complex-visual tasks. These scores contrast with the means for easy-textual and complex-textual tasks at 2.33 and 2.48. Combination and combo-complex tasks also produced moderate scores for task familiarity at 2.69 and 2.67, out of 5.0.

Subjects' assessment of task representation, or how tasks reflected actual tasks facing science educators, revealed other interesting observations. Subjects reported higher scores for combination tasks, which produced an overall mean of 3.93 out of 5.0. Easy (M=3.48) and complex tasks (M=3.67) were also both shown to reflect K – 12 Science Education tasks. In regards to complex tasks, again, distinctions between complex-textual and complex-visual tasks were exhibited. Complex-visual tasks were judged to be significantly more reflective of actual tasks, at a mean of 4.05, than complex-textual tasks, which averaged 3.29 out of 5.0.

Post-search questionnaires also assessed the ease of searching specific task types. These results showed that easy tasks were considered more straightforward, on average, at 4.14 out of 5.0. Moreover, easy-textual and easy-visual tasks both produced identical means at 4.14. Other task types were also shown to be somewhat easy to search including: complex tasks at 3.40, combination tasks at 3.17, and combo-complex tasks at 3.10.

Table 23 also presents subjects' self-assessment of task completion. Subjects felt that easy tasks were completed to a greater extent and, therefore, scored them at a higher average, 4.07 out of 5.0. Subjects also judged other task types to be completed at a moderate to high rate, including mean scores of 3.45 for complex tasks, 3.29 for combination tasks, and 3.12 for combo-complex tasks. Results deriving from this analysis didn’t exhibit any significant differences between the textual and visual forms of these search tasks, in regards to subjects’ scores of task completion.

Results for these task-related factors were also produced across different task types and system variants, collectively (Table 24). Moreover, the results shown in Table 24 depict how scores from the post-search questionnaire varied according to task characteristics and system features.

First, let's examine the composite results for each system variant, as shown in the rows labeled *Overall* (Table 24). Several key observations can be made from these results. When analyzing the scores, or subjects’ response, for each of these task-related factors, System 3 exhibited higher averages. Moreover, mean scores produced by System variant 3 included: 3.12 for subjects’ familiarity with search tasks, 3.84 for task representation, 4.02 for ease of searching individual topics, and, finally, 4.14 for self-assessment of task completion. On the other hand, System variant 2 demonstrated several of the lower averages. System 2 produced means of 2.57 for task familiarity, 3.27 for task representation, 2.61 for ease of searching, and 2.52 for task completion. Averages for System variant 1 were shown to be in-between these high and low scores (please refer to Table 24 for exact results).

Results for System 1, specifically, were used to examine differences among task-related factors during textual, visual, and hybrid search tasks. By surveying results across these tasks categories, several interesting observations can be made. First, visual-only tasks (while using System variant 1) produced higher scores for each of these task-related factors. Results for the visual tasks exhibited that subjects assessed: familiarity with the tasks at 3.79, accurateness of task design at 4.14, ease of searching tasks at 4.29, and overall level of task completion at 4.43. Conversely, text-only tasks produced lower scores for some of these factors, including means of 1.93 for task familiarity, 2.93 for task representation, and 3.43 for task completion.

Subjects' assessment of these task-related factors, for System variant 2, across textual, visual, and hybrid tasks produced results that differed from System 1. For example, text-only tasks led to some of the higher averages. Representation of search tasks, i.e. how each task reflected "real" K – 12 Science Education tasks, measured 3.64 out of 5.0 for textual tasks. In addition, ease of searching textual tasks averaged 3.36 and subject's self-assessment of task completion demonstrated a mean of 3.29. On the other hand, hybrid tasks produced some of the lower scores, including 2.39 for task familiarity, 1.96 for ease of searching, and 1.93 for subjects' assessment task completion.

System variant 3, as employed throughout textual, visual, and hybrid tasks, presented other findings. Visual-only tasks, again, scored some of the higher averages for these task-related factors, such as 3.71 for familiarity with the search tasks and 4.00 for accurateness of task representation. Textual tasks, however, produced high scores for ease of searching and subjects' self-assessment of task completion at 4.29, apiece. When evaluating System 3, during textual, visual, and hybrid task types, results for each of these task-related factors were still considerably high compared to results exhibited by other system variants. For example, although hybrid tasks demonstrated the lowest average for (subjects') self-assessment of task completion and ease of searching, the scores were still significant at 4.04 and 3.89 out of 5.0, respectively.

Table 24: Mean scores for post-search questions relating to search tasks, across different system variants and task types.

System	Task Type	Familiarity with Search Task, Topic *(SD)*	Representation of Search Task *(SD)*	Ease of Searching on Topic *(SD)*	User Assessment of Task Completion *(SD)*
1	**Overall**	2.75 *(1.34)*	3.66 *(1.08)*	3.73 *(1.23)*	3.79 *(1.44)*
	Text-only	1.93 *(1.07)*	2.93 *(1.33)*	3.57 *(1.34)*	3.43 *(1.60)*
	Visual-only	3.79 *(1.31)*	4.14 *(0.86)*	4.29 *(0.83)*	4.43 *(0.76)*
	Hybrid	2.64 *(1.16)*	3.79 *(0.88)*	3.54 *(1.29)*	3.64 *(1.55)*
	Easy	2.71 *(1.77)*	3.21 *(1.25)*	3.64 *(1.39)*	3.64 *(1.39)*
	Easy-Textual	1.43 *(0.79)*	2.71 *(1.38)*	3.00 *(1.53)*	2.86 *(1.57)*
	Easy-Visual	4.00 *(1.53)*	3.71 *(0.95)*	4.29 *(0.95)*	4.43 *(0.79)*
	Complex	3.00 *(1.24)*	3.86 *(1.23)*	4.21 *(0.80)*	4.21 *(0.80)*
	Complex-Textual	2.43 *(1.13)*	3.14 *(1.35)*	4.14 *(0.90)*	4.00 *(1.53)*
	Complex-Visual	3.57 *(1.13)*	4.57 *(0.54)*	4.29 *(0.76)*	4.43 *(0.79)*
	Combination	2.64 *(1.45)*	4.00 *(0.96)*	3.71 *(1.27)*	4.00 *(1.41)*
	Combo-Complex	2.64 *(0.84)*	3.57 *(0.76)*	3.36 *(1.34)*	3.29 *(1.64)*
2	**Overall**	2.57 *(1.11)*	3.27 *(1.20)*	2.61 *(1.37)*	2.52 *(1.55)*
	Text-only	2.50 *(1.23)*	3.64 *(1.22)*	3.36 *(1.78)*	3.29 *(1.73)*
	Visual-only	3.00 *(1.11)*	3.00 *(1.18)*	3.14 *(1.17)*	2.93 *(1.49)*
	Hybrid	2.39 *(1.03)*	3.21 *(1.20)*	1.96 *(0.88)*	1.93 *(1.27)*
	Easy	3.00 *(1.24)*	3.43 *(1.34)*	4.43 *(0.76)*	4.21 *(0.76)*
	Easy-Textual	2.71 *(1.25)*	4.29 *(0.76)*	4.86 *(0.38)*	4.57 *(0.54)*
	Easy-Visual	3.29 *(1.25)*	2.57 *(1.27)*	4.00 *(0.82)*	3.86 *(0.90)*
	Complex	2.50 *(1.09)*	3.21 *(1.12)*	2.07 *(1.00)*	2.00 *(1.00)*
	Complex-Textual	2.29 *(1.25)*	3.00 *(1.29)*	1.86 *(1.22)*	2.00 *(1.53)*
	Complex-Visual	2.71 *(0.95)*	3.43 *(0.98)*	2.29 *(0.76)*	2.00 *(1.41)*
	Combination	2.57 *(0.85)*	3.71 *(0.91)*	1.93 *(0.73)*	1.79 *(1.12)*
	Combo-Complex	2.21 *(1.19)*	2.71 *(1.27)*	2.00 *(1.04)*	2.07 *(1.44)*
3	**Overall**	3.12 *(1.27)*	3.84 *(1.04)*	4.02 *(1.07)*	4.14 *(1.12)*
	Text-only	2.79 *(1.37)*	3.71 *(1.14)*	4.29 *(1.07)*	4.29 *(1.07)*
	Visual-only	3.71 *(1.20)*	4.00 *(1.11)*	4.00 *(1.11)*	4.21 *(0.98)*
	Hybrid	3.00 *(1.19)*	3.82 *(0.98)*	3.89 *(1.07)*	4.04 *(1.23)*
	Easy	3.21 *(1.12)*	3.79 *(1.12)*	4.36 *(0.75)*	4.36 *(0.75)*
	Easy-Textual	2.86 *(0.69)*	3.71 *(1.11)*	4.57 *(0.54)*	4.57 *(0.54)*
	Easy-Visual	3.57 *(1.40)*	3.86 *(1.22)*	4.14 *(0.90)*	4.14 *(0.90)*
	Complex	3.29 *(1.59)*	3.93 *(1.14)*	3.93 *(1.33)*	4.14 *(1.33)*
	Complex-Textual	2.71 *(1.89)*	3.71 *(1.25)*	4.00 *(1.41)*	4.00 *(1.41)*
	Complex-Visual	3.86 *(1.07)*	4.14 *(1.07)*	3.86 *(1.35)*	4.29 *(1.11)*
	Combination	2.86 *(1.23)*	4.07 *(0.83)*	3.86 *(0.95)*	4.07 *(1.14)*
	Combo-Complex	3.14 *(1.17)*	3.57 *(1.09)*	3.93 *(1.21)*	4.00 *(1.36)*

Readers can also begin to examine how subjects responded for each of these task-related factors across different systems and other task categories created for this study. Moreover, Table 24 also reports results for System variants 1, 2, and 3, during easy, complex, combination, and combo-complex tasks.

Results for System 1 revealed a variety of notable observations. For example, subjects' demonstrated a higher familiarity with complex search tasks at a mean at 3.00 out of 5.0. Scores depicting subjects' familiarity with other task types were also somewhat similar, including mean scores of 2.71 for easy tasks and 2.64 for both combination and combo-complex search tasks. In addition, results for easy and complex tasks showed other interesting patterns, especially regarding their textual and visual attributes and influence on subjects' familiarity with search tasks. For example, while subjects' familiarity with easy tasks averaged 2.71 – using System 1 – easy-visual tasks actually demonstrated a score of 4.00, compared to easy-textual tasks at 1.43. Complex tasks, which averaged 3.00 overall for task familiarity, showed similar patterns where complex-visual tasks achieved a significantly higher mean (M= 3.57) than complex-textual tasks (M= 2.43).

Analyzing the accurateness of task design – as also evaluated using System 1 – introduced other findings. Subjects estimated that combination tasks were more reflective of "real" science educational tasks scoring a mean of 4.00 out of 5.0. On the other hand, results for System 1 also showed that easy tasks were deemed less representative, but still relatively positive, at a mean of 3.21. While scores for complex tasks fell in-between easy and combination tasks (M=3.86), again, differences between the textual and visual forms of complex tasks were evident. Moreover, complex-visual tasks were judged to be significantly more accurate, or representative, than complex-textual tasks at a mean of 4.57, compared to 3.14.

Evaluation of System variant 1 also included measuring the ease of searching easy, complex, combination, and combo-complex tasks. Throughout the search experiments, subjects were asked to rate the ease for searching each task with System 1. Readers can observe (Table 24) that complex tasks were deemed easier to search, using System variant 1, at an average of 4.21 out of 5.0. On the other hand, subjects estimated that combo-complex tasks were less easy at a mean score of 3.36. While easy tasks demonstrated an overall mean of 3.64, results for easy-textual and easy-visual tasks varied quite considerably; that is easy-visual tasks produced a mean of 4.29 – in regards to ease of searching – compared to 3.00 for easy-textual tasks.

Readers can also examine (Table 24) how subjects assessed their levels of task completion – as performed on System 1 – across these individual task types. These results indicated that complex tasks produced higher percentages of task completion with a mean of 4.21 out of 5.0[114]. While combo-complex tasks produced lower averages (M=3.29) for this analysis, results were still reasonably significant. Other task types also exhibited moderate to high scores, or averages, including combination and easy tasks, which scored 4.00 and 3.64, respectively. Again, results for the easy tasks revealed distinct differences between easy-visual and easy-textual tasks. Moreover, subjects felt

[114] One possible explanation for such a result could be the exclusion of the Title Browse feature from System variant 1. As previously stated, future analyses will address and justify implications for potential task biases within these results and additional rounds of search experiments.

they performed significantly better on easy-visual tasks, at a mean of 4.43, compared to 2.86 for easy-textual tasks.

This study also examined results for these task-related factors as they scored while evaluating System 2. For example, as part of the post-search questionnaire, subjects were asked to judge their familiarity with easy, complex, combination, and combo-complex search tasks when using System variant 2. Easy tasks produced some of the higher levels for task familiarity at a mean of 3.00 of 5.0, while combo-complex tasks demonstrated a lower average at 2.21. Scores for combination and complex tasks fell in-between easy and combo-complex tasks with averages of 2.57 and 2.50, respectively. No real distinctions between the results for the textual and visual forms of easy and complex tasks, as they influenced subjects' familiarity with the search tasks, were recognized.

Also, from Table 24, readers can observe subjects' reactions to the representation, or accuracy, of easy, complex, combination, and combo-complex tasks while using System variant 2. Results for System 2 showed that combination tasks were judged to be more reflective of K – 12 Science Education tasks, at a mean of 3.71, and that combo-complex tasks were less reflective at 2.71. These results also exhibited that easy tasks were quite representative of "real" tasks at an average of 3.43; however, this analysis revealed differences among easy-textual and easy-visual search tasks. Easy-textual tasks demonstrated a significantly greater average for task accurateness ($M = 4.29$) than easy-visual tasks ($M = 2.57$).

Results collected with System variant 2 showed considerable variation between the ease for searching easy, complex, combination, and combo-complex tasks. Moreover, subjects felt that easy tasks were particularly simple to search – using System

2 – and scored them at an overall mean of 4.43 out of 5.0. On the other hand, combination and combo-complex tasks were deemed significantly more difficult, and only averaged 1.93 and 2.00 out of 5.0, respectively.

Scores also varied when examining subjects' self-assessment of task completion during the evaluation of System variant 2. Moreover, barring easy tasks, which scored a mean of 4.21 out of 5.0, many other task types didn't produce significant results for task completion. For example, combination tasks only averaged 1.79, when estimating task completion ratios, and complex and combo-complex tasks only scored 2.00 and 2.07, respectively. Again, results for System 2 revealed no significant differences among textual and visual task types, in regards to how they influenced subjects' self-assessment of task completion.

Post-search results were finally produced to analyze subjects' perception of these task-related factors while using System variant 3. Similar to the other analyses, described above, subjects' familiarity with the easy, complex, combination, and combo-complex tasks was first examined. Results for System 3 were somewhat similar to those produced for other system variants where complex tasks achieved higher means for task familiarity, at 3.29 out of 5.0, while combination tasks averaged a lower score at 2.86. Subjects were also shown to possess somewhat similar – and significant – levels of familiarity with easy (M=3.21) and combo-complex (M=3.29) search tasks.

Subjects' response to the representation of search tasks, while operating System variant 3, was also quite consistent. Evaluation of System 3 demonstrated that combination tasks were judged to positively reflect K – 12 Science Education tasks at a mean of 4.07 out of 5.0. On the other hand, combo-complex tasks were measured marginally less, but still relatively accurate, at 3.57. Results for the easy and complex tasks fell between with averages of 3.93 and 3.79, respectively.

System variant 3 was also used to measure the ease of searching individual topics. From this analysis, all task types, including easy, complex, combination, and combo-complex tasks, were judged – by the subjects – to be relatively simple to search. Easy tasks scored an overall mean of 4.36, a high among this analysis, while combination tasks scored a low of 3.86. These scores indicated an overall range of means of only 0.50. Complex and combo-complex tasks were also judged to be rather easy to search as each scored 3.93, apiece. Again, results depicting the ease for searching these task types on System 3 did not indicate much difference between the influences of textual and visual forms of search tasks.

Scores depicting subjects' self-assessment of task completion, while using System variant 3, differed from the results produced for System 2. Moreover, analysis of System 2 demonstrated lower averages for task completion across many of these task categories; however, scores produced for System variant 3 were all relatively high. Easy tasks scored higher rates for (subjects') self-assessment of task completion at 4.36 out of 5.0. In addition, these results for easy tasks were comparable to the scores for combo-complex tasks at 4.00. Complex and combination tasks also exhibited significant averages for task completion at 4.14 and 4.07, respectively. These results indicated an overall range of means for subjects' self-assessment of task completion, for System 3, at only 0.36.

The subjective analysis performed for this study also assessed various systems-related factors. These systems-related factors were developed to measure subjects' satisfaction with search results, system functionality, interface support, and the usefulness of certain search and browse features. Results observed for these systems-related factors are summarized in Tables 25 and 26. Similar to other analyses, each of these factors was evaluated compositely and according to different task types and system variants. These systems-related factors were also evaluated using the post-search questionnaire.

Table 25: Mean scores for post-search questions relating to systems-related factors.

Task Type	**Satisfaction with Search Results *(SD)***	**Adequacy of System Functionality for Task Completion (*SD*)**	**Interface Support for Task Completion *(SD)***	**Usefulness of Visual Search Functions *(SD)***	**Usefulness of Keyword Search *(SD)***	**Usefulness of Title Browse *(SD)***
Overall	3.41 *(1.51)*	3.74 *(1.22)*	3.59 *(1.24)*	3.14 *(1.44)*	3.65 *(1.53)*	2.28 *(1.35)*
Text-only	3.60 *(1.52)*	3.88 *(1.27)*	3.67 *(1.30)*	2.26 *(1.33)*	3.71 *(1.55)*	2.74 *(1.52)*
Visual-only	3.69 *(1.30)*	3.90 *(1.08)*	3.83 *(1.03)*	3.79 *(1.34)*	3.68 *(1.51)*	2.14 *(1.29)*
Hybrid	3.18 *(1.58)*	3.58 *(1.24)*	3.43 *(1.28)*	3.25 *(1.34)*	3.61 *(1.54)*	2.11 *(1.25)*
Easy	4.02 *(1.18)*	4.19 *(0.89)*	4.02 *(0.98)*	3.31 *(1.60)*	3.62 *(1.46)*	2.70 *(1.51)*
Easy-Textual	4.00 *(1.45)*	4.33 *(0.97)*	4.05 *(1.12)*	2.24 *(1.34)*	3.62 *(1.43)*	3.32 *(1.57)*
Easy-Visual	4.05 *(0.87)*	4.05 *(0.81)*	4.00 *(0.84)*	4.38 *(1.02)*	3.62 *(1.53)*	2.06 *(1.16)*
Complex	3.26 *(1.52)*	3.60 *(1.35)*	3.48 *(1.29)*	2.74 *(1.42)*	3.77 *(1.59)*	2.21 *(1.34)*
Complex-Textual	3.19 *(1.50)*	3.43 *(1.40)*	3.29 *(1.38)*	2.29 *(1.35)*	3.80 *(1.70)*	2.20 *(1.28)*
Complex-Visual	3.33 *(1.56)*	3.76 *(1.30)*	3.67 *(1.20)*	3.19 *(1.37)*	3.75 *(1.52)*	2.21 *(1.44)*
Combination	3.21 *(1.54)*	3.79 *(1.14)*	3.57 *(1.25)*	3.29 *(1.35)*	3.56 *(1.58)*	2.31 *(1.36)*
Combo-Complex	3.14 *(1.63)*	3.38 *(1.32)*	3.29 *(1.31)*	3.22 *(1.35)*	3.66 *(1.51)*	1.93 *(1.12)*

From examining Table 25, readers can examine the composite results for each of these measures. The composite results, or results across all system variants and task types, can be found (Table 25) in the row indicated *Overall*. Subjects' assessment of many of these systems-related factors produced significant results. For example, the overall mean reported for the effectiveness of system functionality – as it supported task completion – was 3.74 out of 5.0. In addition, subjects also deemed the user interface to positively facilitate task completion at a mean score of 3.59. Composite results were also favorable when evaluating the effectiveness, or accuracy, of search results at 3.41 out of

5.0. These results demonstrated that the experimental systems, including the retrieval, presentation, and interaction with video, were – overall – satisfactory.

Table 25 presents other composite results that reflect subjects' assessment of the usefulness of several interface features and functions. For example, across all task types and system variants, textual searching was determined to be useful at a mean score of 3.65 out of 5.0. Subjects also considered the visual search features useful, or supportive, for task completion, at an average of 3.14. The Title Browse feature was judged least helpful and the scores were more mixed at a mean of 2.28 out of 5.0.

Similar to other analyses performed for this study, these systems-related factors were also evaluated across textual, visual, and hybrid search tasks. Results from this analysis showed that the system's functionality, search results, and interface design positively supported task completion during these different task types. For example, visual-only tasks produced significant scores for each of these different factors, including means of 3.69 for accurateness of search results, 3.90 for adequacy of system functionality, and 3.83 for satisfaction with interface support. On the other hand, hybrid tasks exhibited lower – but still positive – scores for each of these measures; subjects rated their satisfaction with search results at 3.18, the adequacy of system functionality at 3.58, and the level of interface support at 3.43. Mean scores for text-only tasks fell in-between these highs and lows at 3.60, 3.88, and 3.67, respectively. These scores, and the minimal range of means, demonstrated moderate to high satisfaction for each of these systems-related factors.

Readers can also examine (Table 25) subjects' evaluation of individual interface features and functions across textual, visual, and hybrid search tasks. The usefulness of visual searching, not surprisingly, achieved higher scores on visual-only search tasks at a mean of 3.79 out of 5.0. Hybrid tasks exhibited slightly less favorable scores for visual searching with a mean of 3.25. Interestingly, the supportiveness of the visual search features was judged significantly lower for text-only tasks, at an average of only 2.26. Results depicting subjects' assessment of textual searching, across textual, visual, and hybrid tasks, were more consistent. Subjects estimated a mean of 3.71 for keyword searching on textual tasks, which was one of the more significant results from this analysis. However, visual and hybrid tasks scored only slightly lower for textual searching at means of 3.68 and 3.61, respectively. Appraisal of the Title Browse feature was also fairly constant. Subjects regarded the Title Browse as most useful during text-only tasks, at a mean a 2.74, followed by visual tasks, at 2.14, and hybrid tasks at 2.11.

Other task categories, e.g. easy, complex, combination, and combo-complex tasks, were also employed to evaluate these systems-related factors. This analysis assessed subjects' satisfaction with search results, system functionality, and interface support while performing these different task types. Results showed that subjects were more satisfied with the search results, or retrieval, for easy tasks (*M*=4.02). Subjects' satisfaction with search results demonstrated other, more constant, scores across complex, combination, and combo-complex tasks, including means of 3.26, 3.21, and 3.14, respectively. Subjects were also shown to be highly satisfied with system functionality – as it supported the completion of easy tasks – at a mean of 4.19 out 5.0. Scores for system functionality during complex (*M*=3.60), combination (*M*=3.79), and combo-complex tasks (*M*=3.38) were, again, slightly lower, but more constant. The user interface was also deemed significantly more supportive for easy tasks, with a score of

4.02. Other scores depicting subjects' impressions of user interface support included: 3.48 for complex tasks, 3.57 for combination tasks, and 3.29 for combo-complex tasks. Results (Table 25) from this analysis did not demonstrate any significant difference between the textual and visual forms of easy and complex tasks, as they influenced subjects' perceptions of these systems-related factors.

When evaluating subjects' opinions of individual interface features and functions, as they supported the completion of easy, complex, combination, and combo-complex tasks, other interesting trends can be observed. Visual features were deemed significantly more useful for searching easy tasks, where subjects rated their satisfaction at 3.31 out of 5.0. Subjects also believed the visual search features were less useful for completing complex tasks at an average of 2.74. Results for other task types, including combination and combo-complex tasks, scored in-between these scores with averages of 3.29 and 3.22, respectively. Interestingly enough, differences between the textual and visual forms of easy and complex search tasks became apparent when evaluating the usefulness of visual searching. Results showed that subjects deemed visual search features significantly more useful during easy-visual tasks than easy-textual tasks, at a mean of 4.38 compared to 2.24. A similar trend was also found when analyzing results within complex tasks; subjects appraised visual search features to be more useful for complex-visual tasks (M=3.19) than complex-textual tasks (M=2.29).

Subjects' impression of the keyword search, during easy, complex, combination, and combo-complex tasks, revealed other interesting findings. Keyword searching was found relatively useful for complex tasks, at a mean of 3.77, combo-complex tasks, 3.66, and easy tasks, at 3.62. Furthermore, subjects believed that keyword searching was slightly less useful – but still productive – for combination task at 3.56 per task. What's interesting about these results is the differences between the scores for the textual and visual forms of easy and complex search tasks. Contrary to the results for visual searching, subjects' assessment of keyword searching did not vary across the text-only and visual tasks types. Subjects' rated the usefulness of the keyword search feature to be identical across easy-textual and easy-visual tasks at 3.62, apiece. A similar pattern was observed after evaluating the usefulness of the keyword search during complex tasks; complex-textual tasks led to an average of 3.80 while complex-visual tasks produced a mean of 3.75. Results from this analysis demonstrated that the usefulness of keyword searching – as it supported task completion – was quite comparable across these task types.

This study next analyzed the usefulness of the Title Browse feature across easy, complex, combination, and combo-complex tasks. As reported in an earlier discussion, the usefulness of Title Browse feature was not shown to be as significant as the keyword or visual search features, subjectively speaking. Easy tasks produced higher results for the Title Browse feature at 2.70 out of 5.0. The Title Browse feature was deemed even less supportive for combo-complex tasks at an overall mean of 1.93. Results for the easy tasks, again, exhibited differences among easy-textual and easy-visual task types. The Title Browse feature was assessed to be significantly more useful during easy-textual tasks, at a mean score of 3.32, while easy-visual tasks averaged 2.06. On the other hand, complex tasks, which scored an overall mean of 2.21 for the usefulness of the Title Browse feature, demonstrated that results for the complex-textual and complex-visual tasks were mostly similar. Moreover, subjects judged the usefulness of the Title Browse feature at 2.20 out of 5.0 for complex-textual tasks, compared to 2.21 for complex-visual tasks.

The post-search analysis also explored subjects' response to these systems-related factors, not only across the various task types – as previously discussed – but also across different systems. From examining Table 26, readers can first observe the overall results produced by each system variant. The results for System 1 revealed several interesting observations. Subjects assessed the adequacy of system functionality, as it supported task completion, to be significant at a mean of 4.13 out of 5.0. Other systems-related factors were also judged at significant levels for System variant 1, including the supportiveness of the user interface at 3.75 and satisfaction with search results at a mean a 3.66. Subjects' assessment of more specific system components included that the keyword search was judged to be quite useful for completing the various task types at a score of 4.45 out of 5.0. In addition, visual search functions were deemed moderately supportive for task completion, while using System variant 1, and measured 3.27.

Results generated for System variant 2 led to other findings. Overall, the systems-related factors for System 2 produced results that were significantly lower than those achieved by System 1. For example, subjects rated the adequacy of system functionality, as it supported task completion, at 2.91, followed by supportiveness of the user interface, at 2.84, and search results, at 2.48. From examining subjects' perceptions of specific interface features, results showed that the usefulness of the visual search functions averaged 2.82 out of 5.0, while the Title Browse feature, as used on System 2, exhibited a score of 2.22.

Table 26: Mean scores for post-search questions relating to systems-related factors, across different system variants and task types.

System	Task Type	Satisfaction with Search Results *(SD)*	Adequacy of System Function for Task Completion *(SD)*	Interface Support for Task Completion *(SD)*	Usefulness of Visual Search Functions *(SD)*	Usefulness of Keyword Search *(SD)*	Usefulness of Title Browse *(SD)*
1	**Overall**	3.66 *(1.34)*	4.13 *(0.97)*	3.75 *(1.08)*	3.27 *(1.47)*	4.45 *(0.87)*	N/A
	Text-only	3.14 *(1.46)*	4.07 *(1.00)*	3.57 *(1.02)*	2.00 *(1.11)*	4.29 *(1.14)*	N/A
	Visual-only	4.29 *(0.73)*	4.50 *(0.76)*	4.07 *(0.73)*	4.21 *(1.12)*	4.71 *(0.47)*	N/A
	Hybrid	3.61 *(1.42)*	3.96 *(1.04)*	3.68 *(1.25)*	3.43 *(1.37)*	4.39 *(0.88)*	N/A
	Easy	3.50 *(1.61)*	4.07 *(1.07)*	3.43 *(1.02)*	3.36 *(1.74)*	4.21 *(1.12)*	N/A
	Easy-Textual	2.57 *(1.72)*	3.71 *(1.25)*	3.00 *(1.16)*	2.00 *(1.41)*	3.71 *(1.38)*	N/A
	Easy-Visual	4.43 *(0.79)*	4.43 *(0.79)*	3.86 *(0.69)*	4.71 *(0.49)*	4.71 *(0.49)*	N/A
	Complex	3.93 *(0.83)*	4.50 *(0.65)*	4.21 *(0.58)*	2.86 *(1.41)*	4.79 *(0.43)*	N/A
	Complex-Textual	3.71 *(0.95)*	4.43 *(0.54)*	4.14 *(0.38)*	2.00 *(0.82)*	4.86 *(0.38)*	N/A
	Complex-Visual	4.14 *(0.69)*	4.57 *(0.79)*	4.29 *(0.76)*	3.71 *(1.38)*	4.71 *(0.49)*	N/A
	Combination	3.86 *(1.29)*	4.29 *(0.73)*	3.93 *(1.21)*	3.64 *(1.28)*	4.36 *(1.01)*	N/A
	Combo-Complex	3.36 *(1.55)*	3.64 *(1.22)*	3.43 *(1.28)*	3.21 *(1.48)*	4.43 *(0.76)*	N/A
2	**Overall**	2.48 *(1.51)*	2.91 *(1.27)*	2.84 *(1.29)*	2.82 *(1.39)*	N/A	2.22 *(1.32)*
	Text-only	3.29 *(1.73)*	3.43 *(1.60)*	3.36 *(1.69)*	2.36 *(1.34)*	N/A	2.93 *(1.59)*
	Visual-only	2.71 *(1.33)*	3.07 *(1.00)*	3.21 *(1.19)*	3.43 *(1.56)*	N/A	2.15 *(1.28)*
	Hybrid	1.96 *(1.32)*	2.57 *(1.14)*	2.39 *(0.96)*	2.75 *(1.27)*	N/A	1.89 *(1.07)*
	Easy	4.14 *(0.86)*	4.21 *(0.80)*	4.36 *(0.93)*	3.43 *(1.60)*	N/A	3.08 *(1.55)*
	Easy-Textual	4.71 *(0.49)*	4.86 *(0.38)*	4.86 *(0.38)*	2.29 *(1.38)*	N/A	3.57 *(1.62)*
	Easy-Visual	3.57 *(0.79)*	3.57 *(0.54)*	3.86 *(1.07)*	4.57 *(0.79)*	N/A	2.50 *(1.38)*
	Complex	1.86 *(1.17)*	2.29 *(0.99)*	2.21 *(0.98)*	2.36 *(1.28)*	N/A	2.07 *(1.27)*
	Complex-Textual	1.86 *(1.22)*	2.00 *(0.82)*	1.86 *(0.90)*	2.43 *(1.40)*	N/A	2.29 *(1.38)*
	Complex-Visual	1.86 *(1.22)*	2.57 *(1.13)*	2.57 *(0.98)*	2.29 *(1.25)*	N/A	1.86 *(1.22)*
	Combination	1.71 *(0.91)*	2.79 *(1.12)*	2.50 *(0.86)*	2.86 *(1.29)*	N/A	1.79 *(1.12)*
	Combo-Complex	2.21 *(1.63)*	2.36 *(1.15)*	2.29 *(1.07)*	2.64 *(1.28)*	N/A	2.00 *(1.04)*
3	**Overall**	4.09 *(1.18)*	4.18 *(0.94)*	4.18 *(0.92)*	3.33 *(1.43)*	4.34 *(0.98)*	2.42 *(1.46)*
	Text-only	4.36 *(1.08)*	4.14 *(1.10)*	4.07 *(1.07)*	2.43 *(1.56)*	4.29 *(1.27)*	3.07 *(1.44)*
	Visual-only	4.07 *(1.21)*	4.14 *(0.95)*	4.21 *(0.89)*	3.71 *(1.27)*	4.07 *(1.00)*	2.15 *(1.46)*
	Hybrid	3.96 *(1.23)*	4.21 *(0.88)*	4.21 *(0.88)*	3.59 *(1.28)*	4.50 *(0.79)*	2.19 *(1.42)*
	Easy	4.43 *(0.76)*	4.29 *(0.83)*	4.29 *(0.73)*	3.14 *(1.56)*	3.86 *(1.23)*	2.86 *(1.66)*
	Easy-Textual	4.71 *(0.49)*	4.43 *(0.79)*	4.29 *(0.76)*	2.43 *(1.40)*	4.00 *(1.41)*	3.86 *(1.46)*
	Easy-Visual	4.14 *(0.90)*	4.14 *(0.90)*	4.29 *(0.76)*	3.86 *(1.46)*	3.71 *(1.11)*	1.86 *(1.22)*
	Complex	4.00 *(1.41)*	4.00 *(1.18)*	4.00 *(1.18)*	3.00 *(1.57)*	4.50 *(0.94)*	2.38 *(1.33)*
	Complex-Textual	4.00 *(1.41)*	3.86 *(1.35)*	3.86 *(1.35)*	2.43 *(1.81)*	4.57 *(1.13)*	2.29 *(0.95)*
	Complex-Visual	4.00 *(1.53)*	4.14 *(1.07)*	4.14 *(1.07)*	3.57 *(1.13)*	4.43 *(0.79)*	2.50 *(1.76)*
	Combination	4.07 *(1.14)*	4.29 *(0.83)*	4.29 *(0.91)*	3.36 *(1.45)*	4.43 *(0.65)*	2.43 *(1.45)*
	Combo-Complex	3.86 *(1.35)*	4.14 *(0.95)*	4.14 *(0.86)*	3.85 *(1.07)*	4.57 *(0.94)*	1.92 *(1.38)*

Scores for System variant 3 were all relatively significant across these systems-related factors (Table 26). For example, subjects deemed the adequacy of system functionality and usefulness of the interface – as each supported task completion – to both be positive at 4.18 out of 5.0, apiece. Subjects were also pleased with the search results as retrieved by System 3 and estimated their satisfaction at 4.09. System 3, which implemented all experimental features and functions, was also used to assess individual system components. These results showed, again, that subjects were highly satisfied with the supportiveness of the keyword search at a mean of 4.34 out of 5.0. Subjects also rated the visual search features, while using System 3, to be moderately useful at a mean a 3.33 and the Title Browse feature, again, to be somewhat less supportive at 2.42.

Readers can next examine results for these systems-related factors across different task types and system variants, collectively. Moreover, this study explored how subjects assessed these different factors during textual, visual, and hybrid search tasks, while using System variants 1, 2, and 3 (Table 26). Results for System 1 showed that subjects demonstrated higher satisfaction with search results, system functionality, and the user interface during visual-only search tasks[115]. Results varied for the textual tasks, while using System 1, where subjects considered search results significantly less effective (M=3.14) and the user interface not as supportive (M=3.57). Hybrid task, however, showed slightly lower scores for subjects' satisfaction with system functionality, at a mean of 3.96.

[115] Visual-only tasks achieved means of 4.29, 4.50, and 4.07 for each of these systems-related factors, respectively.

Subjects' assessment of more-specific interface features can also be examined (Table 26) using the results for System variant 1. First, visual-only tasks demonstrated significant scores when measuring the supportiveness of certain interface features and functions on System 1. Subjects' satisfaction with the visual search functions, during visual tasks, presented a mean of 4.21. Conversely, the support of visual search functions on textual tasks was measured at only 2.00. Scores for the keyword search feature, on System variant 1, were shown to be more constant; visual-only tasks demonstrated a higher mean at 4.71, followed by textual and hybrid tasks, which produced averages of 4.29 and 4.39, respectively.

Evaluating these systems-related (subjective) factors, across textual, visual, and hybrid search tasks, showed interesting trends for System variant 2. Subjects judged their satisfaction with search results, system functionality, and the user interface to be higher during textual search tasks, at mean scores of 3.29, 3.43, and 3.36, respectively. Averages for these different systems-related factors on System 2 were lower among hybrid tasks, including means of 1.96 for satisfaction with search results, 2.57 for supportiveness of system functionality, and 2.39 for usefulness of the interface. Subjects' assessment of the more individual interface features – as they supported textual, visual, and hybrid tasks – showed other interesting findings. The visual search features on System 2 were judged to be more supportive during visual-only tasks, collectively, at a mean of 3.43[116]. Hybrid and textual search tasks, on the other hand, exhibited lower scores for visual searching at 2.75 and 2.36, respectively. The supportiveness of the Title

[116] Readers should keep in mind that the ViewFinder system performed visual searches using a QBE technique, which required users to find a related (similar) keyframe before searching.

Browse feature on System 2, again, varied by task type, with higher scores observed for textual tasks at a mean of 2.93. Hybrids tasks also produced lower scores, or satisfaction, with the Title Browse feature and averaged 1.89.

Next, readers can observe variations in the post-search results during textual, visual, and hybrid tasks on System variant 3, the full system. Post-search results demonstrated that hybrid tasks led to significant scores for several of these systems-related factors. Moreover, subjects judged that system functionality and the user interface positively supported the completion of hybrid tasks at a score of 4.21, apiece. Subjects estimated that the user interface also facilitated successful completion of visual-only tasks at a rate of 4.21. It was also determined that the search results retrieved for textual tasks were better, or more satisfactory, while using System 3, where subjects assessed a positive average of 4.36. Scores for the search results of hybrid tasks were shown to be slightly lower at a mean of 3.96.

Analysis of more post-search results and specific interface features implemented for System 3 produced other observations. Subjects' satisfaction with the visual search features on visual-only tasks averaged 3.71, while hybrid tasks exhibited a slightly lower mean at 3.59. Text-only tasks demonstrated a significantly lower score for visual searching at 2.43. Results depicting subjects' satisfaction with the keyword search, on System variant 3, were more constant. Subjects rated the keyword search to be supportive for completing hybrid tasks at a mean of 4.50 of 5.0. This mean is comparable with the results for textual and visual search tasks, which scored slightly lower at 4.29 and 4.07, respectively. The usefulness of the Title Browse feature on

System 3 was judged higher for text-only tasks at 3.07, followed by visual-only and hybrid tasks at 2.15 and 2.19.

This study then explored subjects' assessment of these systems-related factors, as they scored for easy, complex, combination, and combo-complex search tasks and different system variants. From this analysis, readers can observe (Table 26) subjects' satisfaction with the search results, system functionality, and user interface on System 1. Results showed that the subjects were mostly satisfied with the search results during complex tasks, at a mean of 3.93, while combo-complex tasks exhibited a lower mean at 3.36. Differences among subjects' assessment of the search results for easy-textual and easy-visual tasks highlight other findings. Moreover, subjects were shown to be significantly more satisfied with the search results for easy-visual tasks, at a mean of 4.43, than easy-textual tasks at 2.57.

Complex tasks also produced higher scores for system functionality with a mean of 4.50. These scores can be compared to the system functionality as judged during combo-complex tasks, which averaged 3.64. Subjects' assessment of the user interface, as it supported task completion, again, showed complex tasks to achieve a higher average at 4.21, while easy and combo-complex tasks measured lower levels at 3.43, apiece. The low to moderate range of means from this analysis, along with moderate to high scores, indicated reasonable and constant effectiveness for these systems-related factors, as they measured for System variant 1.

Next, readers can examine subjects' assessment of some individual interface features and functions on System variant 1 (Table 26). For example, the supportiveness of visual search features was higher for combination tasks, at an average of 3.64 out of

5.0, while complex tasks demonstrated a lower satisfaction at 2.86. Although complex tasks averaged 2.86 and easy tasks averaged 3.36 – when evaluating the visual search features – differences between the textual and visual forms of these task types were evident. Not surprisingly, the visual search features were deemed significantly more supportive during visual tasks, including means of 4.71 and 3.71, compared to 2.00 for the textual tasks.

Some contrasting results for System 1 were uncovered when assessing the usefulness of the keyword search. Keyword searching was judged highly supportive during complex tasks where scores averaged 4.79 out of 5.0. The usefulness of the keyword search on System 1 was slightly less effective for easy tasks at a mean of 4.21. Results for the easy tasks demonstrated marginal differences between easy-visual and easy-textual search tasks. Easy-visual tasks actually produced higher scores for keyword searching (M=4.71) than easy-textual tasks (M=3.71). Averages for combo-complex and combination tasks fell in-between these high and low scores, but still achieved significant means at 4.43 and 4.36, respectively.

These systems-related factors were evaluated – for System variant 2 – across easy, complex, combination, and combo-complex search tasks (see Table 26). Results for System 2 presented interesting variations throughout this analysis. Subjects' satisfaction with the search results of easy tasks was significant at 4.14 out of 5.0. The satisfaction of search results for other task types demonstrated significantly lower averages, including means of 1.71 for combination tasks, 1.86 for complex tasks, and 2.21 for combo-complex tasks. Other differences among subjects' satisfaction with search results were observed during easy tasks. Moreover, subjects deemed that the search results for easy-textual tasks were highly satisfactory, at 4.71, while easy-visual tasks also produced a relatively positive – but significantly lower – score at 3.57.

Distinctions among these task types were also observed when exploring subjects' satisfaction with system functionality on System variant 2. Moreover, easy tasks produced higher rates for system functionality at 4.21 out of 5.0; this result contrasted with scores for complex tasks (M=2.29) and combo-complex task (M=2.36). Again, results for the easy-textual and easy-visual tasks varied where easy-textual tasks achieved nearly a perfect score at 4.86 and easy-visual tasks produced a mean of 3.57.

Subjects' evaluation of the supportiveness of the user interface exhibited similar results. Results indicated that the interface was most supportive for easy tasks, at a mean of 4.36, while also exhibiting differences between easy-textual and easy-visual tasks, with scores of 4.86 and 3.86, respectively. On the other hand, complex tasks produced lower scores, as the user interface was judged to support task completion at 2.21. Combo-complex and combination tasks also led to significantly lower scores when evaluating the user interface and averaged 2.29 and 2.50, respectively.

Subjects' satisfaction with some specific interface features and functions on System 2 revealed other observations. The visual search features were deemed, by the subjects, to be more supportive for completing easy tasks (M=3.43). Scores for other task types were somewhat constant in that visual search features were found to support combo-complex tasks at a rate of 2.64, combination tasks at 2.86, and complex tasks at 2.36. As anticipated, distinctions between easy-textual and easy-visual tasks, in regards to the usefulness of the visual search, were apparent. Moreover, subjects' judged that the visual search features were significantly supportive for easy-visual tasks, at a mean of

4.57, and not as much so for easy-textual tasks, at 2.29. The usefulness of the Title Browse feature also varied according to task type. Easy tasks achieved higher averages for the Title Browse feature, as well, with a mean of 3.08. Easy-textual tasks showed higher levels of satisfaction (*M*=3.57) with the Title Browse than easy-visual tasks (*M*= 2.50).

Next, this study examined subjects' assessment of these factors, as they were scored for the full system variant, or System 3 (Table 26). These results highlighted several observations. Easy tasks, again, produced significant levels for many of these systems-related factors. For example, easy tasks showed higher means for the effectiveness of search results at 4.43 out of 5.0. Conversely, combo-complex tasks averaged a lower, but still relatively productive, score of 3.86 for subjects' satisfaction with search results. The supportiveness of the system's functionality (for System 3) also resulted in significant scores for easy tasks (*M*=4.29). Other task types also scored quite positive, when assessing system functionality, including combination tasks at 4.29, combo-complex tasks at 4.14, and complex tasks at a mean of 4.00. Scores depicting subjects' assessment of the user interface, as it supported task completion, was practically identical to the scores for system functionality; easy and combination tasks produced higher means at 4.29, apiece, followed by combo-complex and complex tasks at 4.14 and 4.00, respectively. As readers can observe, mean scores for these systems-related factors on System 3 contrasted with the results produced by System variant 2, where scores were typically less significant.

Lastly, the post-search analysis of this study examined subjects' satisfaction with individual search features on the full system variant, i.e. System 3. Visual searching was shown to be somewhat supportive across most of these task types. Combo-complex tasks achieved a significant score for visual searching at 3.85, while easy and complex tasks averaged 3.14 and 3.00, respectively. However, differences between the results for visual and textual task types were evident. For example, complex-visual and easy-visual tasks produced scores of 3.57 and 3.86, when assessing the usefulness of the visual search, compared to complex-textual and easy-textual, which measured 2.43, apiece.

Evaluation of the keyword search feature, as it supported task completion, on System 3, also demonstrated positive scores. Keyword searching on combo-complex tasks was shown to be advantageous, and averaged 4.57 out of 5.0. This was closely followed by complex and combination tasks, which also produced significant means at 4.50 and 4.43, respectively. Easy tasks showed a lower, but still relatively positive, score for keyword searching at 3.86. The Title Browse feature was, again, judged to be less productive than either the visual or keyword search features. However, results presented that easy tasks did lead to a higher scores for the Title Browse at a mean of 2.86. Easy-textual and easy-visual tasks demonstrated distinctions among subjects' satisfaction with the Title Browse feature, where easy-textual tasks produced a significantly higher mean (M= 3.86) than easy-visual tasks (M=1.86).

6.3.2 Post-Experiment Results

After each search experiment, other subjective data was collected and evaluated using the post-experiment questionnaire (Appendix P) and short interviews (Appendix Q). These post-experiment methods allowed researchers to appraise a variety of different parameters. Moreover, as part of the post-experiment questionnaire, subjects judged the ease of using and learning to use the ViewFinder system. The post-experiment questionnaire also evaluated ViewFinder's capacity for allowing subjects to assess the nature of video content and relevancy of search results. Subjects were also asked to estimate the necessity for playing the video. The post-experiment questionnaire then had subjects rate the overall effectiveness, or usefulness, of each individual search feature, including the keyword search, Title Browse, attribute weighting, and all visual search capabilities[117]. Finally, the interview session followed up on the post-experiment questionnaire and asked subjects to reflect on the interface features and functions that they deemed useful and ineffective. Results from the post-experiment analysis are summarized in Tables 27 through 36.

First, these post-experiment results observed how subjects assessed some of the more general systems-related factors, or attributes. The composite results presented in Table 27 depict subjects' input across all samples, system variants, and search tasks. Table 27 presents results for the ease of searching and learning to use ViewFinder, and the ability (of subjects) to assess search results and video content. Subjects also estimated the necessity for playing the video content, in order to determine completion of each search task.

Results demonstrated somewhat moderate to high averages for several of these systems-related factors. These results indicated a mean score of 3.62 for the ease of using the system, and 3.92 for the "learnability" of the system. Subjects also deemed that

[117] The post-search questionnaire (Appendix O) collected subjects' opinions about the effectiveness of the Color, Shape, Texture, All-Visuals, Hybrid, and Textual Promote features.

it was relatively easy to determine the relevancy of video content at an average of 3.75 of 5.0. A more moderate score of 2.50 was observed for subjects' perceived necessity of watching the actual video, as it supported task completion.

Table 27: Post-experiment results for systems-related factors.

	Ease of Using System	Ease of Learning System	Ease of Assessing Video Content	Necessity of Playing Video
Total	3.62 *(0.90)*	3.92 *(0.89)*	3.75 *(1.01)*	2.52 *(1.19)*

These general systems-related factors, as evaluated using the post-experiment questionnaire, can also be analyzed according to different subject samples, or sub-samples. First, Table 28 allows readers to examine subjects' estimations for the ease of using and learning to use ViewFinder, the (systems') capacity for allowing assessment of video content, and the necessity for playing the video, each across different age groups. These results indicated that the thirties group deemed the ViewFinder system easier to operate with a score of 4.50 of 5.0. On the other hand, subjects in their fifties thought otherwise and only assessed a score of 2.00 for the ease of using the system. The forties and twenties groups demonstrated moderate to high means for this parameter at 4.14 and 3.31, respectively.

When exploring the "learnability" of the ViewFinder system for different age groups, more consistent results were observed (Table 28). Moreover, subjects in their thirties believed the system was rather easy to learn, and assessed this parameter at 4.50 out of 5.0. The fifties group estimated a lower – but still rather positive – score for learning to use the system at a mean of 3.50. Again, the twenties and forties age groups exhibited moderate to high scores for learning the system at 3.62 and 4.29, respectively.

Evaluating ViewFinder's capacity for allowing the subjects to assess video content – for determining relevancy of search results and task completion – also produced constant results. Subjects in their forties believed the system provided adequate functionality for assessing video, and estimated this quality of ViewFinder at 4.14. Scores were lower for the fifties group, who produced a mean score of 3.33. These scores indicated an overall range of means of only 0.81 across the different age groups.

Table 28: Post-experiment results for systems-related factors by certain age groups.

Age Group	Ease of Using System	Ease of Learning System	Ease of Assessing Video Content	Necessity of Playing Video
20 – 29	3.31 *(0.48)*	3.62 *(1.04)*	3.54 *(1.13)*	2.50 *(1.00)*
30 – 39	4.50 *(0.58)*	4.50 *(0.58)*	4.00 *(1.00)*	1.80 *(1.10)*
40 – 49	4.14 *(0.69)*	4.29 *(0.49)*	4.14 *(0.90)*	3.00 *(1.41)*
50 – 59	2.00 *(1.41)*	3.50 *(0.71)*	3.33 *(0.58)*	2.67 *(1.53)*

Slightly more variation was observed after evaluating subjects' perceived necessity for playing the video in order to determine relevancy (Table 28). Subjects in their forties believed it was more necessary to watch the playing video, at a mean of 3.00, than the thirties group, which assessed it at only 1.80. In addition, mean scores of 2.50 and 2.67 were observed for those in their twenties and fifties.

These systems-related factors were also evaluated among the different grade levels taught, or planned to teach, by the subjects (Table 29). Secondary teachers deemed that ViewFinder was considerably easier to use and learn than the other sub-samples, and estimated the values for these factors at 4.20 and 4.00, respectively. The scores produced by secondary teachers can be compared with elementary and high-school teachers, who assessed the ease of using the system at 3.25 and 3.78. Even more variation was shown among the scores measuring the capabilities for assessing video content and search results. Secondary teachers estimated that the system provided more support for assessing the video at an average of 4.40, while high school and elementary teachers demonstrated moderate scores at 3.78 and 3.50. Subjects' evaluation of the necessity for playing the video exhibited constant scores across the different grade levels being taught. Mean scores produced for this factor included: elementary teachers at 2.46, high-school at 2.56, and secondary teachers at 2.60. These scores indicated an overall range of means of only 0.14.

Table 29: Post-experiment results for systems-related factors by grades taught, or plan to teach.

Grade Levels	Ease of Using System	Ease of Learning System	Ease of Assessing Video Content	Necessity of Playing Video
Elementary	3.25 *(1.06)*	3.83 *(0.84)*	3.50 *(0.94)*	2.46 *(1.05)*
Secondary	4.20 *(0.84)*	4.00 *(1.23)*	4.40 *(0.89)*	2.60 *(1.52)*
High School	3.78 *(0.44)*	4.00 *(0.87)*	3.78 *(1.09)*	2.56 *(1.33)*

Next, Table 30 presents averages for each of the more general systems-related factors, as assessed on the post-experiment questionnaire, across the current occupations[118] held by the subjects. Results from this analysis can be used to examine differences among the results produced by current students and school teachers. For example, each of these two samples scored – for the most part – similarly when evaluating these four different parameters. The scores for these two samples included that students assessed the ease of using ViewFinder at 3.60, while teachers estimated a value of 3.53, a range of means of only 0.07. These groups also reported similar scores

[118] Considering there were only one school administrator and science librarian, an insufficient amount of data was produced for these sub-samples to ensure rigor and warrant valid discussion; however, mean scores for these populations are still reported in Table 30.

for the ease of learning the system, including means of 3.60 (students) and 3.95 (teachers). The ease of assessing the video content was also measured for these different samples, and produced scores for the students at 3.60 and teachers at 3.86. Each group also achieved similar averages when measuring subjects' preference for playing the videos, including students at 2.40 and teachers at 2.55, demonstrating a range of mean of only 0.15.

Table 30: Post-experiment results for systems-related factors by current occupation.

Current Occupation	Ease of Using System	Ease of Learning System	Ease of Assessing Video Content	Necessity of Playing Video
School Administrator, Former Science Teacher	5.00 *(.)*	5.00 *(.)*	3.00 *(.)*	1.00 *(.)*
Student	3.60 *(0.55)*	3.60 *(1.52)*	3.60 *(1.34)*	2.40 *(1.34)*
Teacher	3.53 *(0.96)*	3.95 *(0.71)*	3.86 *(0.96)*	2.55 *(1.15)*
Librarian, Former Science Teacher	4.00 *(.)*	4.00 *(.)*	3.00 *(.)*	4.00 *(.)*

These systems-related factors can also be explored across subjects with different levels of education, including those with Bachelor's degrees (as their highest level) and subjects with a Master's or higher (see Table 31). The ease of using the system was reported, for the most part, similarly across these two subject groups. Subjects who held a Master's degree (or higher) believed the system was easier to use (M=3.80) than those with a Bachelor's (M=3.36). Measuring the "learnability" of the system showed some variation among the results; subjects with a Master's degree deemed that the system was easier to learn, at a mean of 4.20, than those with Bachelor's, who averaged 3.55. Means of 3.82 and 3.64 were reported for the support for assessing the video content by those with a Master's and Bachelor's degree, respectively.

Table 31: Post-experiment results for systems-related factors by level of education achieved.

Highest Degree	Ease of Using System	Ease of Learning System	Ease of Assessing Video Content	Necessity of Playing Video
Bachelors	3.36 *(0.51)*	3.55 *(1.04)*	3.64 *(1.12)*	2.60 *(1.08)*
Masters	3.80 *(1.08)*	4.20 *(0.68)*	3.82 *(0.95)*	2.47 *(1.28)*

Tables 32 through 36 present subjects' assessment of some more specific interface features and functions. As part of the post-experiment questionnaire, subjects were asked to rate the overall effectiveness and support provided by the keyword search, Title Browse, and all variations of the Promote feature. Table 32 displays the composite results depicting subjects' evaluation of these individual interface features and functions. Subjects judged the keyword search as being the most effective, or useful, interface feature at an overall mean of 4.57 out of 5.0. In addition, subjects assessed the overall usefulness provided by the Title Browse and Promote feature to be moderately useful at 3.46, apiece. Among the scores for the individual Promote functions, subjects assessed that the color search was the most effective content-based, or visual, feature at a mean of 3.18. Conversely, the Texture Promote was believed to be significantly less useful, and only produced an average of 2.04. Other scores for the visual search features included the Shape Promote at 2.43, All-Visuals Promote at 2.79, and Hybrid Promote at 2.61. Subjects also assessed the Textual Promote to be somewhat effective with a mean of 3.21. Subjects were less satisfied with the attribute weighting feature, and assessed it at a score of 2.82 out of 5.0.

Table 32: Post-experiment results for interface features and functions.

	Keyword Search	Title Browse	Promote Search	Color	Shape	Texture	All-Visuals	Hybrid	Textual Promote	Attribute Weighting
Total	4.57 *(0.50)*	3.46 *(1.11)*	3.46 *(1.07)*	3.18 *(1.22)*	2.43 *(1.14)*	2.04 *(1.00)*	2.79 *(1.07)*	2.61 *(1.20)*	3.21 *(1.10)*	2.82 *(1.31)*

Experimental results also analyzed other factors across the different age groups (Table 33). The effectiveness of the keyword search was significant for all ages, ranging from 4.71 for the forties group to 4.33 for the fifties group. Assessment of the Title Browse feature showed more variation. The thirties group estimated the usefulness of the Title Browse at a mean of 4.40 out of 5.0, compared to the twenties group, which judged its value at 3.15. Promote searching, in general, also showed some variation among these sub-samples. Subjects in their thirties assessed the Promote Search at higher levels, a mean of 4.00, while those in their fifties estimated its effectiveness at 3.00. Scores for the twenties and forties groups fell in between at 3.15 and 3.86, respectively.

Table 33: Post-experiment results for interface features and functions by certain age groups.

Age Group	Keyword Search	Title Browse	Promote Search	Color	Shape	Texture	All-Visuals	Hybrid	Textual Promote	Attribute Weighting
20 – 29	4.54 *(0.52)*	3.15 *(1.21)*	3.15 *(1.21)*	2.69 *(1.25)*	1.85 *(0.80)*	1.46 *(0.66)*	2.31 *(1.03)*	2.08 *(0.76)*	2.77 *(1.17)*	2.54 *(1.27)*
30 – 39	4.60 *(0.55)*	4.40 *(0.55)*	4.00 *(1.00)*	3.80 *(1.10)*	2.60 *(0.89)*	2.80 *(0.84)*	3.60 *(0.55)*	3.20 *(1.64)*	3.80 *(0.84)*	2.80 *(1.30)*
40 – 49	4.71 *(0.49)*	3.29 *(1.11)*	3.86 *(0.90)*	3.86 *(0.90)*	3.57 *(1.27)*	2.43 *(1.27)*	3.43 *(0.98)*	3.57 *(0.98)*	3.57 *(0.98)*	3.57 *(1.27)*
50 – 59	4.33 *(0.58)*	3.67 *(0.58)*	3.00 *(0.00)*	2.67 *(1.16)*	2.00 *(0.00)*	2.33 *(0.58)*	2.00 *(0.00)*	1.67 *(0.58)*	3.33 *(1.16)*	2.33 *(1.53)*

Examining the scores (see Table 33) for the specific Promote functions across these different ages revealed other findings. For example, the Color Promote was deemed more useful by subjects in their forties and thirties at means of 3.86 and 3.80, respectively. The fifties and twenties groups assessed the Color Promote significantly less useful at 2.67 and 2.69. Other observations were made after evaluating the effectiveness of the Shape, Texture, All-Visuals and Hybrid Promote search functions. Interestingly, subjects in the forties group believed that the Shape search was moderately useful at an average of 3.57, while those in their twenties assessed its effectiveness at only 1.85. In addition, the Texture Promote was judged higher by the thirties group (*M*=2.80) and less useful by those in their twenties (*M*=1.46). Subjects in their thirties also judged the All-Visuals Promote more effective than the other age groups, at an average of 3.60. The forties group deemed hybrid searching moderately useful (*M*=3.57), compared to those in their fifties (*M*=1.67). Results for the Textual Promote were slightly more constant; subjects in their thirties thought this feature was most useful, at a mean of 3.80, while the twenties group deemed it less useful at 2.77. The supportiveness of attribute weighting, as implemented for the Hybrid Promote, actually demonstrated some fairly moderate averages across these different age groups, including the forties group assessing its worth at 3.57 and the fifties group averaging 2.33.

Table 34 presents how other sub-samples, comprising the different grade levels taught, evaluated these individual interface features and functions. Keyword searching was shown, again, to rank higher and more consistently. Readers can observe that secondary teachers rated the effectiveness of the keyword search at 4.80, while the averages produced by high school and elementary teachers closely followed at 4.56 and 4.50, respectively. Results for the Title Browse feature were not as significant; secondary teachers assessed its usefulness at a high of 4.00 and high school teachers rated its value at 2.89. Promote searching, in general, produced somewhat similar averages where scores were also constant across these different sub-samples. Secondary teachers assessed the effectiveness of Promote searching at higher levels, an overall mean of 3.60, while scores for high school and elementary teachers followed, at 3.33 and 3.50. These results demonstrated an overall range of means – for the usefulness of the Promote features – at only 0.27.

Table 34: Post-experiment results for interface features and functions by grades taught, or planned to teach.

Grade Levels	Keyword Search	Title Browse	Promote Search	Color	Shape	Texture	All-Visuals	Hybrid	Textual Promote	Attribute Weighting
Elem.	4.50 *(0.52)*	3.64 *(1.15)*	3.50 *(1.09)*	2.71 *(1.38)*	2.07 *(0.83)*	1.93 *(0.92)*	2.36 *(1.01)*	2.00 *(0.88)*	2.79 *(1.12)*	2.64 *(1.34)*
Second.	4.80 *(0.45)*	4.00 *(1.00)*	3.60 *(0.89)*	3.80 *(0.84)*	3.40 *(1.14)*	3.00 *(1.23)*	3.40 *(1.14)*	3.20 *(1.30)*	3.80 *(0.84)*	3.40 *(1.34)*
High School	4.56 *(0.53)*	2.89 *(0.93)*	3.33 *(1.23)*	3.56 *(0.88)*	2.44 *(1.33)*	1.67 *(0.71)*	3.11 *(0.93)*	3.22 *(1.20)*	3.56 *(1.01)*	2.78 *(1.30)*

Next, readers can examine (Table 34) how subjects evaluated individual Promote search capabilities, including each of the visual search features. Subjects who taught secondary science exhibited higher satisfaction with all of the different visual search features. For example, secondary teachers rated the effectiveness of the Color search at 3.80, which contrasts with elementary teachers who valued it at 2.71. In addition, the Shape and All-Visuals features were also assessed at higher averages by secondary teachers, including a mean of 3.40 out of 5.0, apiece. High school teachers did, however, judge that the Hybrid Promote, i.e. searching via both visual and textual attributes, was more effective at an average of 3.20. The Textual Promote was deemed more useful by secondary educators, at 3.80, and then high school teachers, who rated it at 3.56. The attribute weighting feature was scored more moderately by secondary teachers (M=3.40), and slightly lower for high school (M=2.78) and elementary teachers (M=2.64).

Table 35 (below) presents readers with results for individual interface features and functions across the current occupations held by the subjects. Considering that only one school administrator and one science librarian participated in this study, there is not enough of a sample – or experimental data – to sufficiently compare and analyze these results. (However, scores are still reported for each of these sub-samples in Table 35.) This analysis allows readers to observe certain differences between current Science Education students and teachers. Keyword searching was, again, judged by these groups as consistently effective, including means of 4.40 and 4.67 for the students and teachers, respectively. The Title Browse feature was less useful among these sub-samples where teachers rated it more moderately at 3.62, and students assessed it at 2.80. Results for Promote searching, in general, showed that current teachers valued it more positively, at a mean of 3.71, while students deemed it less beneficial at 2.60.

Table 35: Post-experiment results for interface features and functions by current occupation.

Current Occupation	Keyword Search	Title Browse	Promote Search	Color	Shape	Texture	All-Visuals	Hybrid	Textual Promote	Attribute weighting
School Admin., Former Science Teacher	4.00 *(.)*	4.00 *(.)*	3.00 *(.)*	2.00 *(.)*	2.00 *(.)*	2.00 *(.)*	3.00 *(.)*	1.00 *(.)*	3.00 *(.)*	1.00 *(.)*
Student	4.40 *(0.55)*	2.80 *(0.84)*	2.60 *(1.14)*	3.60 *(0.55)*	1.80 (0.84)	1.60 *(0.89)*	3.00 *(1.00)*	2.80 (1.48)	3.80 *(1.10)*	2.00 *(1.00)*
Teacher	4.67 *(0.48)*	3.62 *(1.16)*	3.71 *(1.01)*	3.14 *(1.35)*	2.62 *(1.20)*	2.14 *(1.06)*	2.76 *(1.14)*	2.62 *(1.16)*	3.14 *(1.11)*	3.10 *(1.30)*
Librarian, Former Science Teacher	4.00 *(.)*	3.00 *(.)*	3.00 *(.)*	3.00 *(.)*	2.00 *(.)*	2.00 *(.)*	2.00 *(.)*	3.00 *(.)*	2.00 *(.)*	3.00 *(.)*

The effectiveness of individual Promote searches, as assessed by students and teachers, are also explored throughout this analysis (Table 35). Results demonstrated that students were moderately satisfied with the Color Promote, and estimated its importance at 3.60. The All-Visuals Promote also demonstrated moderate averages across these sub-samples, where students valued it at 3.00 out of 5.0 and teachers rated it slightly lower at 2.76. Teachers also assessed the Shape Promote somewhat favorably at an average of 2.62 while students deemed it significantly less useful at 1.80. The effectiveness of the Hybrid Promote was judged similarly across these two sub-samples with mean averages of 2.80 (students) and 2.62 (teachers). The Textual Promote did show other positive scores; students rated it at 3.80 and teachers assessed it at 3.14. Attribute weighting achieved other moderate averages for these groups as teachers scored its usefulness at 3.10, and students exhibited some dissatisfaction at 2.00.

Readers can examine how scores for individual system features, as assessed using the post-experiment questionnaire, varied between subjects with different levels of education. Results produced by this analysis are presented in Table 36. Moreover, results were evaluated by subjects holding a Bachelor's degree – as their highest level of education – and those with a Master's (or higher). Subjects' satisfaction with the keyword search, Title Browse, and Promote features were somewhat constant across these two sub-samples. Keyword searching exhibited higher averages, including mean scores 4.55 from those with a Bachelor's and 4.59 according to those with a Master's. Subjects' satisfaction with the Title Browse feature demonstrated more moderate means at 3.45 (Bachelor's) and 3.47 (Master's). Promote searching, in general, produced somewhat similar averages for these two sub-samples, including averages of 3.18 and 3.65.

Table 36 also presents subjects' assessment of individual Promote searches. Color searching – for the most part – was judged moderately useful across both of these groups. Subjects with a Bachelor's degree assessed color searching at 3.00, while those with a Masters exhibited a mean of 3.29. The All-Visuals Promote was deemed the second most effective visual search, according to these two sub-samples, and was valued at 2.73 by those with a Bachelor's and 2.82 by those with a Master's. The usefulness of the Texture Promote was, again, assessed at insignificant levels where subjects with a Master’s estimated its usefulness at 2.29 and those with Bachelor's rated it at 1.64. These two groups also exhibited similar means for hybrid searching, including averages of 2.55 (Bachelor’s) and 2.65 (Master’s). The Textual Promote demonstrated some more positive scores among these sub-samples; subjects with a Bachelor's and Master's rated its effectiveness at 3.18 and 3.24, respectively. The attributes weighting feature achieved identical means of 2.82 across both sub-samples.

Table 36: Post-experiment results for interface features and functions by level of education achieved.

Highest Degree	Keyword Search	Title Browse	Promote Search	Color	Shape	Texture	All-Visuals	Hybrid	Textual Promote	Attribute Weighting
Bachelors	4.55 *(0.52)*	3.45 *(1.04)*	3.18 *(1.25)*	3.00 *(1.27)*	1.91 *(0.83)*	1.64 *(0.81)*	2.73 *(1.01)*	2.55 *(1.04)*	3.18 *(1.08)*	2.82 *(1.17)*
Masters	4.59 *(0.51)*	3.47 *(1.18)*	3.65 *(0.93)*	3.29 *(1.21)*	2.76 *(1.20)*	2.29 *(1.05)*	2.82 *(1.13)*	2.65 *(1.32)*	3.24 *(1.15)*	2.82 *(1.43)*

Lastly, this post-experiment analysis evaluated (Table 37) subjects’ responses from the short interview sessions concluding each search experiment. Throughout the interviews, subjects were asked to identify the system features that they deemed most effective, or supportive, for completing the search tasks, and those they believed to be ineffective. Not surprisingly, results from these interviews exhibited that the first interface feature, i.e. the feature deemed most important, was typically the keyword search. Moreover, 22 out of the 28 subjects specified that the keyword search was the most useful interface feature, or search tool. However, three (3) subjects indicated that the Color Promote was the most effective feature, while another (1) identified the Title Browse. When analyzing the system features that were preferred second and third best, more variety among the responses was observed. Higher frequencies were reported for the Title Browse and Color Promote, including seven (7) and six (6) subjects identifying these features as being second most useful. Overall, a total of 13 subjects identified the Color Promote as an effective interface feature. The Details feature was scored numerous times as a helpful interface feature, including four (4) subjects specifying it as the second most effective tool and five (5) indicating it as their third choice.

Table 37: Summary of interview responses.

Interface Feature	Expressed Like First	Expressed Like Second	Expressed Like Third	Expressed Dislike First	Expressed Dislike Second
Keyword Search	22	–	3	–	–
Title Browse	1	7	2	4	–
More Button	–	–	1	–	–
Clip Details	–	4	5	1	–
Keyframes	–	2	–	2	2
Boolean Search	–	1	–	–	–
Back Button	–	–	–	–	–
Promote Search	–	–	–	–	–
Textual Promote	–	1	1	–	–
Visual Promote	–	2	4	6	1
Color Promote	3	6	4	–	–
Shape Promote	–	1	1	5	–
Texture Promote	–	–	1	2	4
All-Visuals Promote	–	1	–	–	–
Hybrid Promote	–	–	1	–	–
Attribute Weighting	–	–	–	2	1
Total	26 (28)	25 (28)	23 (28)	22 (28)	8 (28)

Table 37 also presents the system features that were disliked, or deemed ineffective, by the subjects. Seven subjects identified visual searching, in general, as being unhelpful at some point of the interview. More specifically, a total of five (5) and six (6) subjects specified that the Shape and Texture Promote features were unsupportive for task completion. Other features that were identified multiple times as unhelpful included: the Title Browse (4), video representation, or the keyframes (4), and attribute weighting (3).

6.4 Examined Relationships

Next, this study examined many associations between the experimental factors. This evaluation included correlation analysis among different parameters and factors, including those related to system effectiveness, task performance, user interaction, search tasks, and subjects' prior knowledge. The analysis presented in this section included correlation tests between factors spanning both objective and subjective evaluation. Significant results are summarized in Tables 38 through 57.

6.4.1 Between Objective Factors

By examining Table 38, readers can observe significant correlations between several objective factors. Moreover, results summarized in Table 38 depict associations between different task types and during-test measures, i.e. user interaction. Factors used to analyze interaction behaviors consisted of the time spent searching each task and the use of individual interface features and functions, such as the keyword search, results browse, Promote search, and visual search.

Table 38: Correlations among task type and user interactions.

	Time Spent	Use of Keyword	Use of More	Use of Back	Use of Promote	Use of Visual Promote
Task Type	r=.335 $p < 0.01$	r=.253 $p <0.01$	r=.196 $p < 0.05$	r=.173 $p < 0.05$	r=.193 $p < 0.05$	r=.173 $p < 0.05$

Correlation analysis between user interaction and task type revealed six significant associations (Table 38). First, task type showed a significant correlation at $p < 0.01$ with the amount time subjects searched each task. Also, the use of the keyword search was significantly correlated at the *0.01* level with task type. Task type was also correlated with browsing search results, i.e. the use of both the More and Back functions, at $p < 0.05$. Promote searching, in general, and the use of a visual Promote search were also correlated with task type, each at the *0.05* level. These associations among task type and user interaction demonstrated certain trends for subjects' use of specific interface features and functions.

Table 39: Correlations among errors and task performance.

	Task Completion
Errors	r=-.397 $p < 0.01$

Table 40: Correlations among errors and user interactions.

	Use of Visual Promote	Use of Color Promote
Errors	r=.372 $p < 0.01$	r=.344 $p < 0.01$

Next, Tables 39, 40, and 41 show how certain objective factors, including during-test measures and task types, were correlated with the number of errors performed by the subjects. Table 39 presents that task completion ratios, as assessed and calculated by the researchers, were significantly correlated at $p < 0.01$ with the amount of errors for each search task. In addition, results from this analysis (Table 40) also portrayed that the use of visual search functions, including the Color Promote, were correlated at the *0.01* significance level with the number of errors. Task types also exhibited a significant correlation with the amount of errors at the $p < 0.05$ level (presented in Table 41).

Table 41: Correlations among errors and task types.

	Task Type
Errors	r=-.175 $p < 0.05$

6.4.2 Between Subjective Factors

This study next explored associations among individual subjective factors. Results from this analysis are presented in Tables 42 through 46. These tests examined correlations between parameters related to subjects' judgments on system effectiveness, self-assessment of task completion, appropriateness of search tasks, and prior knowledge, i.e. familiarity with the search tasks.

Table 42: Correlations among subjects' assessment of system effectiveness.

	Adequacy of System Functionality	**Support of User Interface**
Search Results Satisfaction	r=.763 $p < 0.01$	r=.825 $p < 0.01$

The subjective factors measuring system effectiveness included subjects' assessment of system functionality, supportiveness of the user interface, satisfaction with search results, and task completion. Table 42 presents correlations between different subjective systems-related factors. These results showed that subjects' assessment of system functionality and the supportiveness of the user interface were both significantly correlated at $p < 0.01$ with subjects' satisfaction with search results.

Table 43: Correlations among subjects' assessment of system effectiveness and task familiarity.

	Adequacy of System Functionality	Support of User Interface	Self-Assessed Completion
Task Familiarity	r=.333 $p < 0.01$	r=.403 $p < 0.01$	r=.443 $p < 0.01$

Table 44: Correlations among subjects' assessment of system effectiveness and task characteristics.

	Self-Assessed Completion	Search Results Satisfaction	Adequacy of System Functionality	Support of User Interface
Task Representation	r=.352 $p < 0.01$	–	r=.515 $p < 0.01$	r= .476 $p < 0.01$
Ease of Searching	r=.875 $p < 0.01$	r=.896 $p < 0.01$	r=.830 $p < 0.01$	r=.860 $p < 0.01$

Table 43 presents how various systems- and performance-related factors correlated with subjects' prior knowledge, i.e. their familiarity with the search tasks. From these results, readers can observe that certain factors, including subjects' assessment of system functionality, supportiveness of the user interface, and task completion, were each correlated at *0.01* with task familiarity.

Table 44, however, summarizes a number of other significant correlations. The accurateness of task representation, i.e. how the tasks reflected "real" science educational tasks, exhibited associations with some of these systems- and performance-related factors. Moreover, task representation was significantly correlated with subjects' self-assessment of task completion, adequacy of system functionality, and the supportiveness of the user interface, each at $p < 0.01$. Subjects' estimate of the ease of searching each task was also associated with certain factors, at *0.01*, including the self-assessed task completion ratio, the usefulness of search results, system functionality, and interface support.

Table 45: Correlations among subjects' assessment of task characteristics.

	Ease of Searching
Task Representation	r=.428 $p < 0.01$

Table 46: Correlations among subjects' assessment of task characteristics and task familiarity.

	Task Representation	Ease of Searching
Task Familiarity	r=.428 $p < 0.01$	r=.464 $p < 0.01$

Readers can also examine associations among task-related factors. Table 45 presents how task representation, or subjects' perceptions about the accurateness of the search tasks, was correlated at $p < 0.01$ with the ease of searching. In addition, these task-related factors, including task representation and ease of searching, both produced significant correlations with subjects' familiarity with the search tasks at *0.01* (Table 46).

6.4.3 Between Objective and Subjective Factors

Tables 47 through 57 present some relationships among factors that were included in both objective and subjective analysis. These tests were designed to evaluate associations between many different factors, including those spanning system effectiveness, task performance, prior knowledge, task characteristics, and interaction behaviors.

From this analysis, readers can first observe how factors related to system effectiveness and task performance were correlated with one another. For example, task completion ratios, as objectively measured by during-test analysis, were correlated with a variety of subjective parameters. These results are summarized in Table 47. Readers can observe that task completion ratios measured by the researchers were significantly correlated with subjects' self-assessment of task completion at $p < 0.01$. In addition, subjects' satisfaction with search results, system functionality, and support of the user interface were all correlated with task completion at the *0.01* level. Subjects' perceptions about the usefulness of specific interface features and functions, including the visual and keyword search, were each significantly correlated with task completion at $p < 0.05$ and $p < 0.01$, respectively.

Table 47: Correlations among subjects' assessment of system effectivenss and task performance.

	Search Results Satisfaction	**Adequacy of System Functionality**	**Support of User Interface**	**Self-Assessed Completion**	**Usefulness of Visual Search**	**Usefulness of Keyword**
Task Completion	r=.585 $p < 0.01$	r=.455 $p < 0.01$	r=.579 $p < 0.01$	r=.572 $p < 0.01$	r=.163 $p < 0.05$	r=.466 $p < 0.01$

Next, readers can examine how task performances related to task familiarity, or subjects' prior knowledge. This evaluation, as presented in Table 48, demonstrated that there was indeed a significant correlation between these two factors, at the $p < 0.05$ level.

Table 48: Correlations among task familiarity and task performance.

	Task Completion
Task Familiarity	r=.186 *p < 0.05*

Task-related factors also exhibited associations with other system and performance factors (Tables 49 and 50). As previously discussed, the task-related factors explored throughout this study included: task representation, ease of searching, and task classifications. Table 49, specifically, presents how task completion ratios were correlated with the accuracy of task representation ($p < 0.05$) and the perceived ease of searching each task ($p < 0.01$). In addition, Table 50 shows how task type and subjects' assessment of various systems and performance factors were related. Moreover, results indicated that task type was correlated with subjects' self-assessment of task completion at the *0.01* level. Subjects' satisfaction with search results, system functionality and the user interface were all also significantly correlated with task type at $p < 0.05$ or better.

Table 49: Correlations among subjects' assessment of task characteristics and task performance.

	Task Completion
Task Representation	r=.185 *p < 0.05*
Ease of Searching	r=-.578 *p < 0.01*

Table 50: Correlations among subjects' assessment of system effectiveness and task type.

	Self-Assessed Completion	Search Results Satisfaction	Adequacy of System Functionality	Support of User Interface
Task Type	r=-.215 $p < 0.01$	r=-.206 $p < 0.01$	r=-.204 $p < 0.01$	r=-.184 $p < 0.05$

Table 51 portrays one correlation between different task-related factors. This analysis indicated that task type, or task classification, was correlated at $p < 0.01$ with the difficulty for searching each task.

Table 51: Correlations among subjects' assessment of ease of searching and task type.

	Ease of Searching
Task Type	r=-.283 $p < 0.01$

Next, Tables 52 and 53 present associations between certain task characteristics and user interactions. Moreover, this study analyzed how task representation and ease of searching correlated with the use of specific interface features and functions. Table 52 specifically shows that task representation was significantly correlated with the use of the More ($p < 0.01$) and Promote ($p < 0.05$) features. Table 53, on the other hand, reveals that the ease of searching each task was correlated at the *0.01* level with the use of many interface features and functions, including the keyword search, results browse[119], Promote search, Visual Promote, Textual Promote, and Title Browse.

Table 52: Correlations among subjects' assessment of task representation and user interactions.

	Use of More	Use of Promote
Task Representation	r=-.166 $p < 0.01$	r= -.161 $p < 0.05$

[119] "Results Browse" refers to the use of (both) More and Back buttons.

Table 53: Correlations among subjects' assessment of ease of searching and user interactions.

	Use of Keyword	Use of More	Use of Back	Use of Promote	Use of Visual Promote	Use of Text Promote	Use of Title Browse
Ease of Searching	r=-.386 $p < 0.01$	r=-.385 $p < 0.01$	r= -.250 $p < 0.01$	r= -.372 $p < 0.01$	r= -.302 $p < 0.01$	r= -.260 $p < 0.01$	r=-.518 $p < 0.01$

Readers can next examine how searching behaviors, i.e. the use of specific interface features and functions, were associated with prior knowledge (Table 54). For example, subjects' familiarity with the experimental search tasks was shown to be significantly correlated, at $p < 0.01$, with viewing the clip Details. There was also a significant correlation between task familiarity and using the Promote features, overall and the Textual and Color Promote features individually (see Table 54 for exact scores).

Table 54: Correlations among task familiarity and user interactions.

	Use of Details	Use of Promote	Use of Text Promote	Use of Color Promote
Task Familiarity	r=-.224 $p < 0.01$	r= -.243 $p < 0.05$	r= -.269 $p < 0.01$	r= -.219 $p < 0.01$

The number of errors, as collected from during-test analyses, showed certain relationships with subjective factors. Table 55 presents that the number of errors for each search task was correlated, at the *0.01* significance level, with subjects' satisfaction with search results, systems functionality, and the user interface. In addition, these results showed that the amount of errors was significantly correlated ($p < 0.01$) with the subjects' self-assessment of task completion. Errors (per task) were also correlated with the subjects' prior knowledge, or their familiarity with the search tasks[120]. Subjects' assessment of the ease of searching each task was also related to the number of errors at *0.01* (Table 57).

Table 55: Correlations among errors and subjects' assessment of system effectiveness.

	Search Results Satisfaction	Adequacy of System Functionality	Support of User Interface	Self-Assessed Completion
Errors	r=-.564 $p < 0.01$	r=-.483 $p < 0.01$	r=-.527 $p < 0.01$	r=-.558 $p < 0.01$

Table 56: Correlations among errors and subjects' assessment of task familiarity.

	Errors
Task Familiarity	r= -.256 $p < 0.01$

Table 57: Correlations among errors and subjects' assessment of ease of searching tasks.

	Ease of Searching
Errors	r= -.610 $p < 0.01$

6.5 Results Conclusions

This chapter has discussed many results deriving from both objective and subjective analyses. First, results were presented to depict subjects' demographics and familiarity with various technological skills and educational activities. Demographic information

[120] Table 51 demonstrates that the correlation between the amount of errors and task familiarity is at the $p < 0.01$ significance level.

described the different occupations and grades levels (taught or planned to teach) of the subjects. In addition, subjects demonstrated moderate to high familiarity with several of the inquired skills, including the use of point-and-click interfaces, searching the Web, and preparing Science Education lessons. However, subjects exhibited less familiarity with incorporating video into the classroom and searching online commercial databases and video digital libraries.

Objective results, including during-test measures, were also discussed throughout this chapter. Task completion ratios, completion times, steps and errors, and subjects' interactions, i.e. the use of individual features and functions, were some of the analyzed objective factors. During-test results were presented both compositely and according to specific task types and system variants. Findings from this objective analysis demonstrated certain distinctions among subjects' use of interface features and functions for different search tasks.

Next, this chapter presented results collected through subjective evaluation. Subjective results encompassed subjects' assessment of system effectiveness, task representation, task performance, and the usefulness of individual interface features and functions. These results showed that subjects' perceptions of the keyword search were fairly constant across many task types, while opinions of the visual search and video browse capabilities were somewhat mixed. In addition, subjects' self-assessed task completion scores were comparable to the task completion ratios measured by the researchers. Subjective results were also collected to analyze subjects' familiarity with the different task classifications, which produced higher scores for visual search tasks than either textual or hybrid tasks.

Readers can also observe variations between these subjective and objective results according to the different experimental systems. For example, System 2 produced significantly lower scores – than Systems 1 and 3 – when comparing task completion ratios and various subjective factors.

Finally, this chapter presented correlation analyses between factors spanning objective and subjective parameters. Results from the correlation tests presented significant associations among search tasks, user interactions, and subjects' perceptions of user interfaces, systems functionality and support.

Chapter 7

Discussion and Implications

This chapter discusses several implications resulting from this study. The findings presented in Chapter 6 provided further understand into the initial exploratory research questions (see Chapter 2). To recap, the experimental questions explored throughout this study included:

- What are the important video search tasks for K – 12 Science Education?
- How do search tasks identified for K – 12 Science Education influence users to interact with video retrieval systems, or video digital libraries?
- How can searching behaviors of science educators translate into useful and supportive interface features and functions for video retrieval systems?
- What features best support video searching in a general context and which ones support K – 12 Science Educational tasks, specifically?
- How do domain-specific search tasks and interface features and functions affect system effectiveness?

The evaluation performed throughout this study encompassed both quantitative and qualitative methods (see Chapter 4). This methodology allowed researchers to explore video retrieval and interface development problems from a task- and domain-centric perspective. This chapter presents several conclusions ensuing from this round of search experiments.

7.1 Video Search Tasks for K – 12 Science Education

The nature of this study involved an exploratory and domain-centric approach to researching user interaction with video retrieval systems. Throughout this study, the researchers first needed to evaluate and understand the video search tasks considered to be important for K – 12 Science Education. The researchers drew numerous conclusions from analyzing the experimental results in Chapter 6. Moreover, the post-search analysis provided means for exploring and comparing different classifications of video search tasks developed for Science Education purposes. The post-search questionnaire asked subjects to rate, on a five-point scale, the accuracy of task representation, i.e. how the experimental search tasks reflected "real" Science Education tasks. These results provided researchers with additional understanding of the characteristics of search tasks considered important for K – 12 Science Education.

Subjects' perception of how the search tasks resembled "real" Science Education tasks, referred to as task representation, produced several findings. As presented in Chapter 6, the overall mean score for task representation was 3.59 out of 5.0. In addition, similar means for task representation were observed for textual, visual, and hybrid task types. Results from the post-search analysis demonstrated that textual tasks achieved an overall mean score of 3.43 out of 5.0, compared to visual tasks at 3.61 and hybrid tasks at 3.71. These scores indicated that the overall range of means among textual, visual, and hybrid tasks was only 0.28. Mean comparisons, including independent samples t-tests and Levene's test for the equality of variances, demonstrated equal sampling across these task types, and no significant differences between their respective scores.

There were, however, certain distinctions between the representation levels of easy, complex, combination, and combo-complex tasks. To recap, complex-visual and combination tasks were judged by the subjects to be more representative of "real" tasks facing science educators, with means scores of 4.05 and 3.93, respectively. Conversely, subjects assessed complex-textual and combo-complex tasks to be less representative of K – 12 Science Education tasks at 3.29, apiece. Mean analyses indicated equal variances and significant differences (at 0.05) among each unique comparison of these high and low averages.

When understanding the search tasks that are important for K – 12 Science Education, it's also important to take a closer look at the tasks that the subjects, i.e. domain professionals, were most familiarity with. Results from this analysis revealed some variations among subjects' familiarity with the different task types. Overall, subjects demonstrated the most familiarity with visual tasks at an average of 3.50 of 5.0. Subjects' familiarity with visual tasks can be compared to the scores for hybrid and textual tasks, which produced means of 2.68 and 2.40, respectively. These averages for visual, textual, and hybrid task types did exhibit significant differences (at 0.05) and equal variances. In addition, differences in subjects' familiarity with easy, complex, combination, and combo-complex tasks were discovered as well. Results, again, showed that subjects were more familiar with the visual forms of these task types, including easy-visual and complex-visual tasks, at means of 3.62 and 3.38, respectively. These scores were found to be significantly different than the averages for easy-textual (M= 2.33) and complex-textual tasks (M=2.48). As previously mentioned, statistical differences were measured using independent samples t-tests and Levene's equality of variances.

From these results, researchers were capable of drawing several conclusions about the search tasks involved in K – 12 Science Education. The researchers concluded that a wide range of video search tasks were equally important for the everyday activities of Science Educators. Overall, i.e. across all search runs, results indicated that textual, visual, and hybrid search tasks were all significant for the domain of Science Education. As previously mentioned, subjects reported no significant differences between the representation level of visual, hybrid, and textual search tasks. Although there were some distinctions among the representation of other task categories, e.g. easy, complex, combination, and combo-complex tasks, no concrete trends were established. For example, by examining the representational scores of these task types, readers can observe that the tasks deemed most reflective of "real" tasks, i.e. complex-visual and combination tasks, both collectively comprised textual and visual attributes, and also involved multiple steps, or sub-tasks. In addition, the tasks that were shown to be less representative of "real" tasks, i.e. complex-textual and combo-complex tasks, also encompassed textual and visual elements, and required multiples of steps as well. As a result, these observations led researchers to conclude that a wide range of task characteristics and information attributes are important for searching Science Education video.

Another significant finding from these results was that not only textual, or semantic, needs were found important for K – 12 Science Education. For example, subjects exhibited higher familiarity with visual tasks either than textual or hybrid tasks. The visual forms of easy and complex tasks also demonstrated higher familiarity than the text-only task types. While it's important to note that search tasks involving specific (and lesson-dependent) information were shown to be significant for K – 12 Science Education, visually-oriented information tended to be equally – if not more – valuable for Science Educators.

So, what contributed to having observed a wide assortment of tasks considered important for K – 12 Science Education? The experimental tasks developed, and found significant, for this study were comprised of needs that ranged from general-visual information to more specific – and lesson-related – semantic information. The researchers believe that the wide assortment of important tasks for K – 12 Science Education is a product of the domain being highly specialized and wide-spread. Moreover, teachers are trained and employed to teach a variety of specific subjects, ranging from cell biology to weather and climate. As a result, the evaluation of search tasks from a more general perspective – symbolic of this study – produced moderate to high scores for task familiarity and representation across many different classifications and needs.

7.2 Task Influence on User Interaction

As previously discussed, the researchers formulated several conclusions about the search tasks considered to be important for Science Education. Subsequently, it's important to begin analyzing how user interaction and searching behaviors were influenced by these experimental search tasks. During-test results, or the use of the individual interface features and functions, were presented in Chapter 6. It’s important for this study to interpret trends in searching behaviors across different task types in order to understand influences and implications for user interactions.

First, let's analyze variations in the during-test results, or searching behaviors, compositely and across different task types. As previously mentioned, keyword searching was performed more than any other action, at a mean of 2.57 times per task. Viewing clip Details (M=2.08), browsing search results (M=1.82) and video titles (M=1.60), and Promote searching (M=1.20) each achieved lower results. However, by delving deeper into these results, readers can begin to observe trends in user interaction across different task types. For example, the results (Chapter 6) make apparent that subjects were more inclined to browse (and re-browse) search results when there were visual characteristics comprising the information need, or search task. In addition, visual searching, in general, and color searching – the most utilized visual search – were also performed more frequently for tasks containing visual attributes. Conversely, keyword searching, video (title) browsing, and viewing clip details, for the most part, remained steady across many task categories.

Other differences in subjects' interactions were found dependent upon the easy, complex, combination, or combo-complex task categories. For example, the rate of keyword searching was significantly different for combination tasks than the complex-visual, complex-textual, and combo-complex search tasks. Browsing search results, again, demonstrated higher frequencies among the tasks with visual qualities. The use of the Promote features, including visual searches, also varied across these different task categories; complex-visual, combo-complex and combination tasks exhibited higher rates of visual searching than easy-textual tasks. In addition, use of the Textual Promote differed across task types, including higher frequencies on complex-textual and combo-complex tasks. Browsing video titles also varied, especially for combination and easy-textual tasks.

The during-test results depicting subjects' interactions, or searching behaviors, are further supported by certain findings from correlation analyses (discussed in Chapter 6). Moreover, correlation tests revealed various associations, at a minimum of $p < 0.05$, between task types and factors related to subject interaction, including completion time(s), keyword searching, results browsing, Promote and visual searching. In addition, factors encompassing task representation and ease of searching also revealed significant correlations with a variety of searching behaviors.

So, the researchers considered: what are the implications for these results and associations? From examining the means, variances, and significant correlations, the researchers can assume – for this particular study – that the subjects did interact with the experimental systems differently during different task types. Moreover, many of the task types explored by this study, including easy, complex, combination, and combo-complex tasks, were considered important for researching the interaction behaviors of Science Educators for domain-specific tasks. In addition, there was also significant variation among the searching behaviors exhibited for textual, visual, and hybrid search tasks. These results supported one initial premise of this study that the developmental process of video search tools should take into consideration the characteristics of domain and search tasks, which were shown to influence user interaction. As a result, the design of user interfaces and specific features and functions can benefit from employing domain- and task-centric methods.

7.3 Implications for Interface Design

The previous section presented certain findings and observations associated with subjects' interaction behaviors. Next, it's important for this study to follow-up on these findings and ask, how do these results affect interface designs for video retrieval systems? Another question deriving from these results included, how can researchers translate significant results into developing useful user interfaces, i.e. features and functions, which support Science Education and video searching in general? Drawing upon the findings and observations of this study, the researchers were capable of making several conclusions.

When examining the interface features and functions designed to support both Science Education and general searching, readers can first observe certain trends within the during-test, post-search, and post-experiment (subjective) results. For example, from examining the post-search data, it is apparent that the keyword search was subjectively judged to be the most supportive feature for task completion, and positively assessed at a mean of 3.65 out of 5.0. Alternatively, subjects rated the usefulness of the visual search features, also from the post-search questionnaire, to support task completion at a level of 3.14 out of 5.0.

However, before drawing many conclusions about the "correct" designs of video retrieval interfaces, it's important to fully understand and recap subjects' assessment of the experimental system features during different task types. For example, the keyword search feature was judged most useful for completing complex-textual tasks, at a mean of 3.80 out of 5.0. On the other hand, the keyword search feature was estimated least helpful during combination tasks, at a mean a 3.56. Subjects' assessment of the visual search features varied more, where visual searching was shown most useful for easy-visual tasks (M=4.38) and least helpful for easy-textual tasks (M=2.24). The Title Browse feature also demonstrated contrasting results, including a high of 3.32 for easy-textual tasks and a low of 1.93 for combo-complex tasks. The (subjective) usefulness of these various interface features and functions was also compared for textual, visual, and hybrid search tasks. These results exhibited similar trends, in that the supportiveness of the keyword search feature was again judged more consistently across different task types, with an overall range of means of only 0.10. When assessing the usefulness of the visual search and Title Browse features, results achieved an overall range of means of 1.53 and 0.63, respectively. (Refer to Chapter 6 for exact scores.)

Results from the post-experiment analysis helped validate some of these observations. Moreover, according to the post-experimental results, the keyword search was again judged to be the most useful feature, at a mean of 4.57 out of 5.0. On the other hand, the Title Browse and Promote features exhibited overall means for their perceived usefulness at 3.46, apiece. The Color Promote was shown to be the most supportive visual feature, at a mean a 3.18, compared to lower means being observed for the Texture and Shape features at scores of 2.04 and 2.43, respectively. In addition, responses during the short interviews also supported the results from the post-experiment questionnaires. Moreover, 22 out of the 26 responding subjects indicated that the keyword search was the most supportive interface feature, while 3 subjects specified the Color Promote, and 1 recognized the Title Browse. When asked to identify the feature that was next (second) most supportive for task completion, some of the visual searches produced further positive reactions, including a total of 10 (out of 25) subjects having recognized one of the visual search capabilities. Also, a total of seven subjects specified the Title Browse as their second favorite. Chapter 6 then presented certain correlations among factors encompassing subjects' assessment of the experimental systems and their prior knowledge. For example, task familiarity was shown to be significantly correlated at 0.05 with subjects' satisfaction with systems functionality and interface support.

So, what are some conclusions about domain-centric interface designs that can be drawn from this evaluation? First, the researchers can safely infer that a keyword search is a valuable interface feature for K – 12 Science Educational video systems. (However, the purpose of this particular study was never to discount the usefulness of a keyword search feature.) The researchers took a closer inspection at the results of the keyword search and developed other, more specific, conclusions about its usefulness. The assessment of the keyword search feature, both objectively and subjectively, produced positive and constant results across all experimental task types. As a result, the researchers concluded that a keyword search feature was applicable in a more general and widespread video retrieval context. Also, it was important that the researchers were able to conclude that the evaluated form of keyword searching, i.e. a transcript-based feature, was highly useful and effective for many task types when implemented properly. These results are significant because the researchers have isolated and tested an individual keyword search feature and produced significant results for a specific professional domain, one advancement from previous video IR studies. The other form of textual searching evaluated throughout this study, i.e. the Textual Promote feature, exhibited a significant difference from the transcript search feature, as subjects preferred formulating their own queries rather querying-by example (QBE). However, the Textual Promote did show some positive results on certain task types and system variants, particularly when the transcript search feature was not available.

Another significant contribution of this study was that certain visual features were in-fact shown to positively support K – 12 Science Education video searching. The benefits observed for these features were significant because many previous video retrieval studies sweepingly dismissed the implications of visual searching as negative or marginal, at best. However, previous video IR research also did not elaborate or closely inspect why visual searching remained ineffective, nor did they explore possible contexts (i.e. domains or search tasks) where visual features could be employed more effectively. These findings highlight the advantages for researching video IR problems from a domain- and task-centric perspective, where researchers were capable of examining the application and benefits for such features.

From examining the composite, or overall, results for the visual search features, readers may conclude that the impact of each was somewhat trivial. However, this notion may be based on the general, or widespread, use of the visual search features being significantly less than keyword searching. The during-test measures and especially the subjective post-search results indicated that the application of the visual search features was more specialized, meaning that the use and implementation of such features was task dependent. These observations contrasted with the results of the keyword search feature, which was shown to be more universally applicable. These results were also compounded by the subjective results, where a wide range of means for visual searching was observed across different task types. This conclusion was then validated by the post-experimental interviews and subjects' recognition of visual search features as helpful across certain task types.

In addition, this study took a closer examination at the results for the visual search features, and evaluated certain attributes for retrieving K – 12 Science Education video. From analyzing the post-experimental results, i.e. questionnaires and interviews, it became apparent that the Color and, quite possibly, the All-Visuals Promote features were somewhat supportive for completing certain task types. On the other hand, based on the during-test and post-search (subjective) analyses, the researchers concluded that the Texture, Shape, and Hybrid features were ineffective, or unsupportive. The results of this study also helped researchers determine that user feedback, or the query weighting feature, was also insignificant for Science Education search tasks.

Next, this study can also assume that the interface features that allowed users to inspect the videos' content and search results were also important. Results from analyzing these features were more mixed, as browsing search results was shown to be applicable in a general sense and also exhibited trends for specific task characteristics. The results browse functions were more useful on tasks that contained visual attributes, especially visual-only tasks. On the other hand, fewer trends – across task types – were observed among subjects' use of the Details feature. Subjects did, however, indicate some inclination for viewing clip details for more conceptually-oriented search tasks.

To recap, there were numerous implications for developing user interfaces of Science Education video search systems. The researchers were satisfied to discover that a transcript-based (keyword) search feature was significant across both subjective and objective analyses. Another important observation of this study provided further understanding about the application and employment of visual search features, including specific visual attributes deemed useful and ineffective for this particular domain. The results of this study also provided researchers with a better understanding of displaying search results and other video information to end users, and how searching behaviors were influenced by the characteristics of domain-specific search tasks.

The researchers of this study believe that some of these patterns were influenced by subjects' familiarity, or their prior exposure and preference with preexisting search tools. Most people are accustomed to searching for information by keyword, such as in using Google. Moreover, keyword searching is already embedded into many people's everyday work-related activities, while visual searching, on the other hand, remains foreign to them. The fact that the experimental results actually demonstrated even a slight significance for visual searching within certain contexts – or different task types – brought about one accomplishment of this study. Results did indicate that subjects' prior knowledge was associated with their assumptions of system functionality and interface features and functions. As a result, the researchers believe that given more exposure and training, visual search and browse features, when implemented specifically for K – 12 Science Education, could achieve even greater significance for everyday searching. Also, future rounds of search experiments may also produce better results for visual search features. Moreover, if the researchers were able to modify and evaluate these interface features and functions over different intervals of time and contexts using the same subject pool, a greater significance may be observed for visual searching in K – 12 Science Education. However, the researchers also believe that the use of visual search features will continue to be more task dependent than keyword searching, and that transcript-based search features, when implemented appropriately, will remain a significant, if not the predominant, search tool for video retrieval.

7.4 Discussion of System Effectiveness

The results analysis performed for this study also allowed researchers to closely examine how search tasks and certain interface features and functions, together, influenced system effectiveness. The different objective factors used to gauge system effectiveness included task completion ratios, completion times, and the numbers of steps and errors. On the other hand, results from this study also assessed system effectiveness by examining subjects' satisfaction with search results, system functionality and interface support, as each were measured to facilitate task completion.

From examining these particular questions, it was important to understand how effective each system variant performed overall. The composite results measuring system effectiveness made apparent that Systems 1[121] and 3[122] achieved significantly greater task completion scores than System variant 2[123]. Furthermore, Systems 1 and 3 achieved completion ratios of 84% and 89%, respectively, while variant 2 demonstrated a percentage of 52%. Subjective scores evaluating system effectiveness reflected similar patterns. Systems 1 and 3 exhibited significantly higher averages for subjects' self-assessment of task completion and satisfaction with search results and system functionality.

The results, as just described, included composite scores for each system variant. From this analysis, the usefulness of certain interface features was shown to be more task dependent than others. Therefore, it was important to analyze how effectively individual system variants performed during different task types. Readers can observe how System variant 1 produced positive completion scores, times, and steps and errors during hybrid and complex tasks. While System variant 2 exhibited fewer significant scores, overall, the researchers did observe encouraging results on certain task types. Moreover, visual tasks – while using System 2 – exhibited a task completion ratio of 85.71%, an average completion time of 3:09, and a mean number of steps and errors at 8.00 and 2.57, respectively. More specifically, easy tasks, including easy-visual tasks, achieved productive scores for System 2. The full system, or System variant 3, produced

[121] System 1 included the keyword and visual search capabilities, but excluded the Title Browse.

[122] System 3 was the full system variant and included the keyword search, full range of visual and Promote capabilities, and the Title Browse.

[123] System 2 included the Title Browse and visual search capabilities, but excluded the keyword search.

significant results for both textual and combination tasks; these task types demonstrated completion ratios at 92.86% and 96.43%. Times for task completion and the numbers of steps and errors across textual and combination tasks were also found to be productive on System 3.

Results differed somewhat when subjectively evaluating system effectiveness according to different task types. System 1 was judged, by the subjects, to be more useful during visual and complex-visual tasks, where satisfaction with search results achieved means scores of 4.29 and 4.14 out of 5.0. In addition, subjects felt that the functionality of System 1 adequately supported the completion of visual (M=4.50) and complex-visual search tasks (M=4.57). The interface of System 1 was also deemed supportive for these particular task types at averages of 4.07 (visual) and 4.29 (complex-visual). System variant 2, which produced lower averages overall, exhibited positive results for easy-textual tasks. Subjects estimated their satisfaction with the search results (on System 2) of easy-textual tasks at 4.71 out of 5.0, while also assessing the supportiveness of the system's functionality and user interface at 4.86, apiece. System 3 was judged most useful during easy and text-only search tasks. Subjects felt System 3 produced positive search results for text-only tasks at 4.36. System functionality and interface support on System variant 3 were also deemed productive for textual tasks at 4.14 and 4.07, apiece.

When examining how task characteristics in conjunction with interface features and functions influenced the effectiveness of the video retrieval system, the researchers also drew upon some observations from the correlation tests (Chapter 6). To recap, the researchers discovered that significant associations did exist between certain task attributes and factors related to system effectiveness. Task type, representation, and ease of searching were also found to be correlated at $p < 0.01$ with subjects' appraisal of system functionality, support of the user interface, and self-assessment of task completion. In addition, subjects' prior knowledge, i.e. task familiarity, was significantly correlated with each of these systems-related factors, each at the 0.01 level. These results demonstrated additional associations among search tasks, interface features, and system effectiveness.

What do these results tell us about video retrieval systems and the influences of search tasks and interface features and functions? First, results from this analysis emphasized the importance of evaluating system features across different task types. Composite results for system effectiveness underscored how incorporating a keyword search, such as in Systems 1 and 3, could produce positive results. Subsequently, findings from this analysis also validated other observations, as previously described in this chapter, including that the transcript-based (keyword) search feature was effective when searching video across different task types. While composite results generated for System variant 2 were somewhat unsubstantial, certain task types exhibited otherwise. For example, Systems 1 and 3 were shown to be more effective during hybrid and textually-oriented search tasks, while System 2 produced positive scores for visual and easy search tasks. Results of the visual and easy tasks supported previous claims that the features implemented for System variant 2, i.e. the visual search and Title Browse features, were more task dependent than a keyword search. In addition, researchers were, again, able to validate that system effectiveness could be improved by incorporating video search features that extend beyond a keyword, or transcript-based, search. Based on objective and subjective analyses, this study concludes that interface developers and researchers need to analyze potential use(s), audience(s), domain(s), and search tasks prior to designing video retrieval systems. Video retrieval systems designed for more general use are likely to benefit from search features that vary from those implemented for specific domains and tasks. Future studies should strive to understand how and when to incorporate different types of video search and browse capabilities.

7.5 Conclusions

This chapter has indicated several benefits for exploring video retrieval research and interface designs from a domain-centric perspective. Results from this study have provided greater insight into experimental parameters and factors frequently overlooked throughout a majority of preexisting video IR research. One notable contribution of this study was that researchers were provided with an understanding of domain-centric search tasks, including how textual, visual, and hybrid video search tasks were all deemed important and prevalent in Science Education. In addition, results from this study also demonstrated how search tasks were essential for exploring video retrieval problems, as tasks exhibited associations and influences with user interaction and system effectiveness. Such associations and influences were discovered using objective and subjective analyses.

From evaluating the interface features and functions of ViewFinder, several different analyses indicated that a keyword (transcript) search feature was important for retrieving K – 12 Science Education video. However, another positive outcome of this study was that significant results were also achieved for various visual search and video browse features. While employing a keyword search was shown to be consistent and effective across different task types, the usefulness of the visual search and video browse features were far more task dependent. Results of the individual interface features and functions also indicated the importance for user- and task-centric evaluation, as the methods employed by this study provided researchers with means for determining effective video search tools for a particular domain. Moreover, evaluation of the different interface features and functions verified that it was inherent for researchers to examine specific tasks, audience(s), and application(s) of video systems. These findings also signify how future studies, which adopt similar approaches for evaluation and development, can further facilitate the creation and validation of standards for designing user interfaces of video retrieval systems geared toward specific domains and tasks.

Chapter 8

Conclusions and Future Developments

This study was developed to explore how domain-centric search tasks influence user interaction with video retrieval systems, and how results can be used to design transparent user interfaces. The underlying goal for this study is to promote alternative means for evaluating video retrieval systems in the hopes that future video IR research will progress towards general interaction models and interface development frameworks. User- and task-centric efforts in video IR are necessary because current evaluation methods are primarily technologically driven and results are showing little progress.

The problems presented throughout this study pertain to multiple areas. First, previous studies that explored and formulated information seeking models are considered important. Several models that depict textual and visual information seeking were presented (Chapter 3). However, research related to visual information seeking is limited. The lack of progress in developing visual models is significant, because theoretical findings can provide support for creating information tools and understanding user interaction. The need to further associate information seeking and information retrieval research has been expressed by other notable researchers (Vakkari, 1999).

The development of user interfaces for video retrieval systems encompasses another area of research that is important for this study. User interface research, as discussed throughout this paper, detailed the implementation of features and functions that support searching and browsing digital video. A recent survey reviewed specific interface features and functions developed by certain research groups, spanning both academia and industry (Lee & Smeaton, 2004). The survey by Lee and Smeaton also presented details about interface implementation and functionality, including query modeling, data representation, and systems evaluation.

Lee and Smeaton (2004) also presented a framework of user interfaces for video retrieval systems; their findings were based on traditional, or textual, information seeking models. Although Lee and Smeaton's findings are significant, further progress is needed because user- and task-centric efforts in video IR research should be explored more independently, i.e. away from text-based studies. Frameworks for developing user interfaces of video retrieval systems should be based on the evaluation of visual information seeking behaviors. Additional progress in user- and task-centric research can be promoted by exploring the users, visual needs, and behaviors exhibited throughout certain domains.

The third area of research considered important to this study involved search tasks for video retrieval. When investigating domain-specific problems in video retrieval, it is imperative that researchers also explore search tasks. Implications of search tasks have been comprehensively examined throughout text IR research, and moderately examined throughout image studies. Video retrieval has neglected task-centric research, and search tasks are widely dismissed when designing user interfaces. Considerations for search tasks in current video IR efforts basically include analyzing the needs for one specific purpose, and developing solutions based solely for that particular project. While a project-specific approach may be beneficial for individualized efforts, current research is not facilitating the creation of standardized tools or interface designs. Considering that search tasks have not been thoroughly evaluated throughout video IR research, future efforts will not be provided any support for assessing user interaction or designing user interfaces. As a result, task-centric research in video IR needs to progress.

One objective for this study was to explore systematic methods for developing and evaluating domain-specific interfaces of video retrieval systems. The domain involved in this study was K – 12 Science Education. Search tasks reflecting the information needs and activities facing science educators, e.g. preparing lab assignments, demonstrations, or lecture materials, were created and evaluated. It is believed that tasks involving Science Education are supported with digital video, and developing means to retrieve video for science educational purposes is an important objective for the field.

The problems raised by this investigation were thoroughly explored. A formal experimental study was carried out by the researchers. The methodology of this study required the researchers to implement several system variants, and to isolate and evaluate the effectiveness of specific interface features and functions. The experimental design involved a total of 28 science educators, including teachers and current majors (students). Each subject performed a total of six search tasks using different system variants. Experimental data was collected using both quantitative and qualitative methods, and evaluation included objective and subjective analysis. Data analysis explored various experimental parameters, including task performances, user interactions, system effectiveness, and task representation, across different system variants and task types.

Results from these experiments provided insight into several experimental parameters frequently overlooked in video IR studies. Researchers were given understanding of domain-centric search tasks, including how textual, visual, and hybrid search tasks were all deemed important by science educators. Results from this study also revealed associations among search tasks, user interaction and systems effectiveness.

Evaluating individual interface features and functions exhibited that keyword searching was significant for retrieving NASA (Science Education) video. On the other hand, another significant finding from this study included that positive results were actually observed for various visual search features. Moreover, results from the keyword search were shown to be more consistent and effective across different task types, while the visual search features were task dependent. These results demonstrate the importance for user- and task-centric evaluation, as they provide researchers with additional understanding for developing useful video search tool.

Future Advancements

The information explosion that has occurred over the past 10 to 15 years will likely motivate other future advancements in video retrieval. Similar to other types of information, i.e. text, audio, and image, digital video is being produced at a remarkable rate[124] (Lyman & Varian, 2003). Film and television are almost exclusively produced in digital form and older, i.e. pre-digital, video content is constantly being digitized.

[124] Annual production of video content, if stored digitally, is estimated in the millions of terabytes.

Evidence of the explosion in digital video can be found by analyzing video digital libraries on the Web[125], many of which host a wide-range of videos that predate digital technology.

The sudden increase in digital video should help advance both technological tools and user-centered research. While this study emphasizes the importance of user-centered research, and specifically explores problems from the perspective of search tasks, it's still beneficial to consider the technological tools which may appear in the near future. Technological advances will facilitate newer and improved methods for video processing, data representation, and information management.

Improvements in video formatting should support automatic representation and indexing. MPEG-4 and MPEG-7, two different video formats, can effectively embed and layer other types of metadata throughout digital video. MPEG-4 will enable "object-based retrieval" by facilitating automatic segmentation and tracking (Smeaton, 2004). MPEG-7 will stream other portable information along with video, similar to that of XML, which can be used to describe the visual, audio, and textual attributes of a particular video or segment (Smeaton, 2004). Although MPEG-4 and MPEG-7 are currently in practice, they are not reaching their full potential.

Researchers should also expect the development of new tools for manually segmenting, annotating, and indexing digital video (Marchionini & Geisler, 2002). Future annotation tools, similar to IBM's VideoAnnEx, will help streamline manual processing, and provide the accuracy of expert analysis. While annotation tools will

---

[125] Some notable online video libraries include the Internet Archive, Open Video (UNC Chapel Hill), and Informedia Project (Carnegie Mellon University).

benefit information professionals, such as librarians, the manual processing of video information will continue to be a laborious and time-consuming task.

Other technological standards for retrieving and managing visual information may also develop. Visual querying languages may be developed to exclusively retrieve multimedia documents. A visual query language may closely resemble some of today's standards, such as SQL, except the application would support retrieving video and image information, as opposed to text. Some current database tools, such as Oracle's interMedia, are already making strides in managing multimedia documents,

Alongside these advances in video indexing and processing, an even greater emphasis will be placed on developing interface tools to effectively support users and tasks. Moreover, with the increasing amount of digital video, many video digital libraries, or other video retrieval systems, may become even more specialized, and users may find it increasingly difficult to find relevant information. Studies that explore problems from a user-centered perspective will be significant because researchers will need to discover which features and functions best support domain-specific tasks. One example of how interfaces are expected to evolve includes video retrieval capabilities being integrated into virtual workspaces, so users can collectively study, analyze, and discuss video content (Marchionini & Geisler, 2002). In addition, researchers must recognize the shift toward mobile technologies, where users will be capable of accessing and consuming video content through portable devices (Smeaton, 2004). Each of these developments in video retrieval will require more analysis of user interactions with video applications.

Once user interfaces for video retrieval systems begin demonstrating more powerful features and functions, developmental frameworks are likely to emerge. Video IR research should be capable of generating some general ideas, or foundations, for developing interfaces to support users' assumptions about digital video and search tasks. As a result, researchers will need a better understanding of the interaction behaviors of users while searching and browsing digital video. Understandings of user behavior can be produced, or derived from, comprehensive user and task analyses, which demonstrate the factors, decisions, and stages of the video seeking process.

It is also believed that user and task analyses will lead to the discovery of search tools, or interface features, that can effectively support specific domains. As a result, video IR researchers can expect the formulation of domain-specific principles which may derive from task-centric research. These principles will be significant because researchers will be provided ideas for developing features and functions for video library projects that are domain-specific in nature. Domain-specific principles will hold interoperable qualities as they can be potentially shared and integrated across other interrelated efforts.

Understanding user interaction with video retrieval systems will not only produce more effective user interfaces, but theoretical works on video information seeking should surface in the more distant future. Researchers may observe the eventual recognition of a video information seeking model. As previously stated, it would be realistic if a video information seeking model resembles a more traditional model that was developed for text-based retrieval; however, a visual model would describe the processes associated with finding video information specifically. Conniss et al. (2000) formulated a general interaction model for describing image seeking behaviors across a variety of domains (presented in Chapter 3). However, future visual models may incorporate other domains, enhance understanding of behaviors across different user groups, and contrast between conceptual and specific searching, text-based and visual retrieval, and a variety of other factors (Conniss et al., 2000). As you can image, formulating a video information seeking model will be a challenging task for future research.

Finally, standards for evaluating video retrieval studies should also continue to develop. Common means for evaluating video retrieval systems will be important, because differing studies will need the capability to cross-compare experimental results. Various conferences and forums, including the TREC Video Retrieval Evaluation Workshop (TRECVID), have begun developing guidelines for evaluating retrieval systems. The challenges associated with common evaluation are extensive, as researchers must adhere to a common protocol, search tasks, dataset, and video segmentation. While progress has been made in places such as TRECVID, more research and progress is needed. Standards to evaluate more interactive experimental studies, including studies that are more user- and task-centric in nature, will prove to be even more challenging.

Bibliography

EVIA Digital Archive - Ethnomusicological Video for Instruction and Analysis. Retrieved June 27, 2005 from http://www.indiana.edu/~eviada

IBM Research - VideoAnnEx Annotation Tool. Retrieved April 15, 2004 from http://www.research.ibm.com/VideoAnnEx/

interMedia. Retrieved April 14, 2007 from

http://www.oracle.com/technology/products/intermedia/index.html

TREC-9 interactive searching study. Retrieved April 8, 2004 from http://www-nlpir.nist.gov/projects/t9i/qforms.html

TRECVID 2003 Guidelines. Retrieved April 7, 2004 from http://www.itl.nist.gov/iaui/894.02/projects/tv2003/tv2003.html

TREC Video Retrieval Evaluation Homepage. Retrieved May 11, 2005 from http://www-nlpir.nist.gov/projects/trecvid/

TRECVID interactive experimental design. Retrieved April 14, 2004 from http://www-nlpir.nist.gov/projects/tv2003/TRECVID_Interactive_Experimental_Design.html

WebSEEK: Content-based Image and Video Search and Catalog Tool for the Web. Retrieved April 7, 2004 from http://www.ctr.columbia.edu/webseek/

Adams, B., Amir, A., Dorai, C., Ghosa, S., Iyengar, G., Jaimes, A. et al. (2003). IBM Research TREC-2002 Video Retrieval System. *Proceedings of the Eleventh Text REtrieval Conference, TREC 2002*.

Amir, A., Berg, M., Chang, S. F., Hsu, W., Iyengar, G., Lin, C. Y. et al. (2004). IBM research TRECVID-2003 video retrieval system. *Proceedings of the Text REtrieval Conference's Video Workshop (TRECVID-2003)*.

Amir, A., Srinivasan, S., & Efrat, A. (2003). Search the audio, browse the video - A generic paradigm for video collections. *Eurasip Journal on Applied Signal Processing*, *2003*(2), 209-222.

Bates, M. J. (1989). The design of browsing and berry picking techniques for the online search interface. *Online Review*, *13*(5), 407-424.

Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, *5*, 133-143.

Belkin, N., Seeger, H., & Wersig, G. (1983). Distributed expert problem treatment as a model for information system analysis and design. *Journal of Information Science, 5*(5), 153-167.

Browne, P., Czirjek, C., Gaughan, G., Gurrin, C., Jones, G. J. F., Lee, H. et al. (2004). Dublin City University video track experiments for TREC 2003. *Proceedings of the Text REtrieval Conference's Video Workshop (TRECVID-2003).*

Budzik, J., & Hammond, K. J. (2000). User interactions with everyday applications as context for just-in-time information access. *Proceedings of the 5th International Conference on Intelligent User Interfaces.*

Choi, Y., & Rasmussen, E. M. (2003). Searching for images: The analysis of users' queries for image retrieval in American history. *Journal of the American Society for Information Science and Technology, 56*(6), 497-510.

Chua, T.-S., Neo, S.-Y., Li, K.-Y., Wang, G., Shi, R., Zhao, M., & Xu, H. (2005). TRECVID 2004 search and feature extraction task by NUS PRIS. *Proceedings of the Text REtrieval Conference's Video Workshop (TRECVID-2004).*

Conniss, L. R., Ashford, A. J., & Graham, M. E. (2000). Information Seeking Behaviour in Image Retrieval: Visor I Final Report.

Cooke, E., Ferguson, P., Gaughan, G., Gurrin, C., Jones, G. L. F., Le Borgue, H., Lee, H., Marlow, S., McDonald, K., McHugh, M., Murphy, N., O'Connor, N. E., O'Hare, N., Rothwell, S., Smeaton, A. F., & Wilkins, P. (2005). TRECVID 2004 experiments in Dublin City University. *Proceedings of the Text REtrieval Conference's Video Workshop (TRECVID-2004).*

Del Bimbo, A. (1999). Visual information retrieval. San Francisco: Morgan Kaufmann Publishers.

Dervin, B. (1977). Useful theory for librarianship: Communication, not information. *Drexel Library Quarterly, 13*(3). 16-32.

Dervin, B., & Nilan, M. (1986). Information needs and uses. *Annual Review of Information Science and Technology, 21*, 3-33.

Eidenberger, H., & Breiteneder, C. (2003). VIZIR - a framework for visual information retrieval. *Journal of Visual Languages and Computing, 14*(5), 443-469.

Garber, S. R., & Grunes, M. B. (1992). The art of search: A study of art directors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*

Gaughan. G., Smeaton, A. F., Gurrin, C., Lee, H., & Mc Donald, K. (2003). Design, implementation and testing of an interactive video retrieval system. *MIR 2003 - 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*.

Goodrum, A., & Spink, A. (2001). Image searching on the Excite web search engine. *Information Processing & Management*, *37*(2), 295-311.

Greisdorf, H. (2002). Information seeking behaviour in image retrieval: Visor I final report. *Journal of Documentation, 58*(3), 346-348.

Hauptman, A., Baron, R. V., Chen, M. Y., Christel, M., Duygulu, P., Huang, C. et al. (2004). Informedia at TRECVID 2003: Analyzing and searching broadcast news video. *Proceedings of the Text REtrieval Conference's Video Workshop (TRECVID-2003)*.

Hauptmann, A., Chen, M.-Y., Christel, M., Huang, C., Lin. W.-H., Ng, T., Papernick, N., Velivelli, A., Yang, J., Yan, R., Yang, H., & Wactlar, H. D. (2005). Confounded expectations: Informedia at TRECVID 2004. *Proceedings of the Text REtrieval Conference's Video Workshop (TRECVID-2004)*.

Hearst, M. (1999). Chapter 10. User interfaces and visualization. *Modern information retrieval* (pp. 257-323).

Hjelsvold, R., Lagorgen, S., Midtstraum, R., & Sandsta, O. (1995). Integrated video archive tools. *Proceedings of the 3rd ACM International Conference on Multimedia (MM '95)*.

Jörgensen, C. (1996). Indexing images: Testing an image description template. *Proceedings of the ASIS Annual Conference*.

Jörgensen, C., & Srihari, R. (1999). Creating a Web-based image database for benchmarking image retrieval systems. *Proceedings of the SPIE - The International Society for Optical Engineering, vol.3644*, 534-541.

Kuhlthau, C. (1988). Longitudinal case studies of the information search process of users in libraries. *Library and Information Science Research, 10,* 257-304.

Lee, H., & Smeaton, A. F. (2004). Designing the user interface for the Físchlár digital video library. *Journal of Digital Information, 2*(4).

Lyman, P., & Varian, H. R. (2003). *How much information*. Retrieved January 23, 2004 from http://www.sims.berkeley.edu/research/projects/how-much-info-2003/

Marchionini, G. (1995). Information seeking in electronic environments. Cambridge*:* Cambridge University Press.

Marchionini, G., & Geisler, G. (2002). The Open Video digital library. *D-Lib Magazine, 8*(12).

McDonald, S., & Tait, J. (2003). Search strategies in content-based image retrieval. *Proceedings of ACM SIGIR '03.*

Norman, D. A. (1990). The design of everyday things. New York: Doubleday.

Saracevic, T. (1996). Relevance reconsidered '96. *Information science: integration in perspective* (pp. 201-218).

Shneiderman, B., Byrd, D., & Croft, B. (1997). Clarifying search: a user-interface framework for text searches. *D-Lib Magazine*.

Simonnot, B., & Smaï, M. Model for interactive retrieval of videos and still images. *Proceedings from the International Workshop on Multi-Media Database Management Systems*.

Smeaton, A. F. (2004). Indexing, browsing, and searching of digital video. *Annual Review of Information Science and Technology* (38 ed., pp. 371-407).

Smeaton, A. F., Kraaij, W., & Over, P. (2004). TRECVID 2003 - An introduction. *Proceedings of the Text REtrieval Conference's Video Workshop (TRECVID-2003)*.

Smeaton, A. F., & Over, P. (2003). The TREC-2003 video track report. *Proceedings of the Eleventh Text REtrieval Conference, TREC 2002*.

Smith, J. R., & Chang, S. (1997). Visually searching the web for content. *IEEE Multimedia, 4*(3), 12-20.

Spink, A., Goodrum, A., & Hurson, A. R. (2001). Multimedia web queries: Implications for design. *Proceedings of the International Conference on Information Technology: Coding and Computing* (p. 589 – 593).

Toms, E. G., Freund, L., Kopak, R., & Bartlett, J. C. (2003). The effect of task domain on search. *Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative Research*.

Vakkari, P. (1999). Task complexity, problem structure and information actions integrating studies on information seeking and retrieval. *Information Processing and Management*, *35*(1999), 819-837.

Vorhees, E. M., & Harman, D. K. (eds.). Common evaluation measures. *NIST Special Publication 500-250: The Tenth Text REtrieval Conference*.

Wildemuth, B. M., Yang, M., Hughes, A., Gruss, R., Geisler, G., & Marchionini, G. (2004). Access via features versus access via transcripts: User performance and satisfaction. *Proceedings of the Text REtrieval Conference's Video Workshop (TRECVID-2003)*.

Yan, R., Yang, J., & Hauptmann, A. G. (2004). Learning query-class dependent weights in automatic video retrieval. *Proceedings of ACM Multimedia (MM'04)*.

Yang, H., Chaisorn, L., Zhao, Y., Neo, S.-Y., & Chua, T.-S. (2003). VideoQA: Question answering on news video. *Proceedings of ACM Multimedia (MM'03)*.

Yang, M., Wildemuth, B., & Marchionini, G. (2004). The relative effectiveness of concept-based versus content-based video retrieval. *Proceedings of ACM Multimedia (MM'04)*.

Zhou, X. S., & Huang, T. S. (2002). Unifying keywords and visual contents in image retrieval. *IEEE Multimedia*.

Appendix A

Email to SLIS-L for Recruiting Users for Pilot Study

Dear SLIS Students,

Do you have a science background? Are you interested in science librarianship? If so, you are needed for a research project that explores how science educators search for and use digital video. This will not be an intrusive experimental study; we are primarily interested in designing user interfaces for video retrieval systems, not evaluating subjects. Subjects should expect to spend 20 minutes for each experimental session. (Subjects may only perform one full experimental session.) During each session, subjects will be asked to perform a total of 5 video search tasks and fill out a pre-test, post-search, and post-experiment questionnaire, and answer several interview questions. Students will be compensated $5 for their time. This study is chaired by Dr. Javed Mostafa (jm@indiana.edu), Associate Dean for Research at SLIS. This study attempts to benefit science education and video retrieval research by designing and evaluating domain-specific video retrieval systems. Search tasks will involve searching and browsing the NASA K - 16 Science Education Programs, an open source video collection. These experiments are part of a dissertation research project.

Please read further details found on Call for Research Participants posted at:
http://ella.slis.indiana.edu/~daalbert/phd/research/

Thanks you very much for your time. Please contact me directly at daalbert@indiana.edu to participate. Your help will be greatly appreciated.

Appendix B

Email to Instructors of Science Education Courses

Dear Professor [   ],

My name is Dan Albertson, a doctoral candidate in the School of Library and Information Science (SLIS). My dissertation research explores how K – 16 science educators, including those in training and current professionals, search for and use digital video for instructional purposes. Findings from this study will be measured as implications for designing interfaces for video retrieval systems. This study will attempt to benefit K – 16 science education and video retrieval research. To explore these issues, I am indexing NASA K-16 Science Education Programs, an open source video collection, and implementing different prototype systems. Using these tools, experiments will be designed to target and measure the effectiveness of domain-specific interface features and functions. I am contacting you because you currently teach science education. I have been involved in video retrieval research for over three years, and am now preparing these experiments for my dissertation research.

I would like to speak with science education students, instructors, and professionals. This will not be an intrusive experimental study; we are primarily interested in designing user interfaces for video retrieval systems, not evaluating subjects. This study is chaired by Dr. Javed Mostafa, Associate Dean for Research at SLIS.

Call for research participants posted at:
http://ella.slis.indiana.edu/~daalbert/phd/research/

Would you be willing to help me out with this study? Basically, I need your help in assembling a subject pool. I have come up with several options on how to reach students:

- I could give a short introduction (5 min) to this project at the end of a class session and pass around a sign-up sheet to those interested in participating. If the class size is too large for such an option, I could provide the URL (above) after my presentation and interested students can contact me directly. By having my presentation at the end of a class session, students who are not interested in participating will be allowed to leave the classroom. You will receive a script of the presentation in advance.

- If there is an email listserv established for your course or science education students, you could provide me access to post a message asking for participants. I would forward the web URL (above) for experimental details.

- I could provide you with the URL (above) to the project webpage so you can pass it along to your students and inform those who are interested to contact me directly. You could notify them of the compensation for participating in the study.

Any or all of these options are acceptable.

Subjects should expect to spend [20 minutes (pilot)] [1 1/2 to 2 hours (formal)] for each experimental session. Experiments will require each subject to perform [5 (pilot)] [12 (formal)] video search tasks and completing a pre-test, post-search, and post-experiment questionnaire, and a short interview. Students will be compensated [$5 (pilot)] [$10 / hour (formal)] cash for their time.
For personal homepage see:
http://ella.slis.indiana.edu/~daalbert/homepage/

Thanks you very much for your time. I look forward to hearing back from you.

Appendix C

Informed Consent Statement for Pilot Study

You are invited to participate in a research study. The purpose of this study is to evaluate how search tasks and professional domain influence interaction with and effectiveness of video retrieval systems. Results of this experimental study will be analyzed as implications for designing user interfaces of video retrieval systems. These experiments are part of a preliminary – or pilot – study.

INFORMATION
Participation in this study requires a number of different tasks including:

1. Fill out a pre-test questionnaire covering background and demographics.
2. Search for 5 assigned search topics, or information needs, using the given video retrieval system. Subjects are encouraged to end each search task whenever he/she feels that successful results have been retrieved. Subjects may discontinue any search task at any time. Behaviors and comments about system usefulness will be noted by the experimenter.
3. Fill out a brief post-search questionnaire after the completion of each search task.
4. Fill out a post-experiment questionnaire after completing all search tasks. The post-experiment questionnaire will ask questions related to the usefulness and effectiveness of the video retrieval system.
5. Answer several questions orally, i.e. a short interview, where responses will be audio recorded.

Audio recordings of post-experiment interviews will be transcribed, coded, and erased within two months of experiment completion. Recordings of the interviews are strictly for research purposes and only the investigators will have access to the content. Audio recordings will not include any reference to the subject's real name or personal identification numbers. Subjects will be allowed to review to their recording at anytime after the conclusion of the experiment. If any subject withdraws from the experiment prior to completion, all audio recordings will be erased immediately. Interviewing questions will focus on the effectiveness of the retrieval system and user interface, not on the subjects themselves.

This experimental session will last for 20 minutes and subjects will be given 3 minutes per search task. This study will involve approximately 5 to 10 human subjects.

RISKS

There are no foreseeable risks associated with this research.

BENEFITS

This study strives to advance video retrieval research. One benefit is that this study investigates problems surrounding video retrieval from a user-centered perspective, whereas the current focus of video retrieval research is predominately systems-centered. This study also attempts to discover why current video retrieval experiments do not produce significant results by incorporating different methods for systems evaluation.

CONFIDENTIALITY

Subject confidentiality will be protected. At no point in the experiments, questionnaires, or interview will real names or personal identification numbers be recorded. All ensuing reports will discuss subjects using an assigned identifier. Experimental data, including audio tapes, completed questionnaires, interview responses, and experimental notes, will be locked in a secure location where only the investigators will be given access. Identity of the subject can only be broken if the subjects choose to discuss their participation with others. Audio recordings deriving from this pilot study will be destroyed by January 1, 2007.

COMPENSATION

For completing this pilot study, i.e. all 20 minutes, you will receive $5 cash. If you withdraw from the study prior to its completion, you will receive prorated compensation based on the $15 per hour and the total time spent performing experiments. For example, if you complete 10 minutes of experiments you will be compensated 15 ($/hour) * 1/6 (hour) = $2.50.

CONTACT

If you have questions at any time about the study or the procedures, you may contact the researcher, Dan Edward Albertson II, at Herman B. Wells Library, LI011, 812-360-0579, and daalbert@indiana.edu.

If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact the office for the Indiana University Bloomington Human Subjects Committee, Carmichael Center L03, 530 E. Kirkwood Ave., Bloomington, IN 47408, 812-855-3067, or by e-mail at iub_hsc@indiana.edu.

PARTICIPATION
Your participation in this study is voluntary; you may refuse to participate without penalty. If you decide to participate, you may withdraw from the study at anytime without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed.

CONSENT
I have read this form and received a copy of it. I have had all my questions answered to my satisfaction. I agree to take part in this study.

Subject's signature____ Date ____________

For further information regarding language used for consent, please see the human subject instruction packet at: http://research.iu.edu/rschcomp/pdf/forms.pdf

All guidelines for research compliance developed by the Office of the Vice President for Research at Indiana University, http://research.indiana.edu/

Further information regarding Protection of Human Subjects at Indiana University, Bloomington can be directly accessed at: http://research.iu.edu/rschcomp/hmpg.html

Appendix D

Search Tasks of Pilot Study

Searcher #___________
Condition ___________

SEARCH TASKS OF PILOT STUDY

1. Find the *Good Stress (GS)* segment *Bones*, part of the *NASA Connect* series. Display the corresponding keyframes. Display the details for the 17th keyframe, i.e. search result.

2. Find two different programs (or segments) that discuss 'Aerodynamics' and also shows images/visuals of the Wright Brothers flight at Kitty Hawk.

3. Find a program that discusses how meteorologists use satellite technology to predict weather forecasts and then find a student activity that simulates cloud types to estimate precipitation.

4. Find shots of an astronaut preparing for or performing a space walk and then shots showing the surface of mars up-close.

5. Find shots that contain visuals of Aurora Borealis being bright blue skies during night. Next, find an activity that examines the potential of other planets experiencing the same phenomena.

Appendix E

Pre-Test Questionnaire of Pilot Study

Searcher #___________
Condition ___________

PRE-TEST QUESTIONNAIRE OF PILOT STUDY

What is your highest level of education? Degree_______ Major__________ Year ____

Are you currently a student? □ YES □ NO If yes, what is your current major and concentration(s)? ____

Do you have any training in Science Librarianship or Science Education? □ YES □ NO If so, which one______________

Are you interested in Science Librarianship or Science Education? □ YES □ NO If so, which one ________________

Do you have a Science background ? □ YES □ NO If so, briefly describe how (i.e. work, schooling, other training, etc.) ____

What is your current occupation? ____

What is your gender? □ FEMALE □ MALE What is your age? ________ years

Please circle the number closest to your experience

How much experience have you had:	None		Some		A great deal
1. Using a point-and-click interface (e.g., Macintosh, Windows)	1	2	3	4	5
2. Searching commercial online systems or databases (e.g. Dialog, Lexis-Nexis, EBSCO)	1	2	3	4	5
3. Searching WWW search engines (e.g., Google, Yahoo)	1	2	3	4	5
4. Conducting science education or science librarianship research	1	2	3	4	5
5. Searching video retrieval systems or video digital libraries	1	2	3	4	5
6. Searching on other systems, please specify the system: ____	1	2	3	4	5

Appendix F

Post-Search Questionnaire of Pilot Study

Searcher #___________
Condition ___________
Topic # _____________

	Not at all		Somewhat		Extremely
1. Are you familiar with the subject matter of this topic?	1	2	3	4	5
2. Are you satisfied with your search results?	1	2	3	4	5
3. Did the system provide adequate functionality to complete this search task?	1	2	3	4	5
4. How representative do you feel this task is of "real" search tasks science instructors or librarians encounter?	1	2	3	4	5

Appendix G

Post-Experiment Questionnaire of Pilot Study

Searcher #___________
Condition ___________

Please consider the searching experience that you just had.

	Not at all		Somewhat		Extremely
1. How easy was it to use this information system?	1	2	3	4	5
2. How easy was it to learn to use this information system?	1	2	3	4	5

Appendix H

Interview Form of Pilot Study

Searcher #___________
Condition ___________

EXPERIMENTER INTERVIEW QUESTIONS AND NOTE FORM OF PILOT STUDY

What comments do you have about interface features and functions?

What specific interface features were most useful and why?

What did you dislike about any of the interface features? Which features did you find confusing and why?

How could the interface be more effective?

Are there any interface features that are missing? Any features you would like to see when retrieving video?

Appendix I

Video Collection Guide for Pilot Study

NASA Video Collection Guide of Pilot Study.
ViewFinder Categories (Acronyms of Full Programs)

NASA CONNECT SERIES

AATC – Ahead Above the Clouds
AO – Ancient Observatories
ATE – A-Train Express
BHFSTE – Better Health from Space to Earth
DITNS – Dancing in the Night Sky
EOM – Eyes Over Mars
FoF – Festival of Flight
FOFE – Future of Flight Equation
GoE – Geometry of Exploration
GS – Good Stress

GWTF – Glow With the Flow
HASB – Having a Solar Blast
HT – Hidden Treasures
ISS – International Space Station
MMOU – Mirror, Mirror On the Universe
MOAT – The Measurement of All Things
MTF – Modeling the Future
PSA – Personal Satellite Assistant
PW – Plane Weather
QTS – Quieting the Skies
RFTF – Recipes for the Future
RTTS – Rocket to the Stars
SOF – Shapes of Flight
TOAT – Tools of the Aeronautic Trade
TRROR – The Right Ratio Of Rest
TWM – The Wright Math
VE – Virtual Earth
VT – Venus Transit
WYGTYA – Wherever You Go There You Are
WFS – Wired For Space
XPG – X-Plane Generation

OTHER SERIES – NO ACRONYMS
Destination Tomorrow
NASA Kids Science News
NASASciFiles
NASAWhy?Files

Appendix J

Results of Pilot Study

J.1 User Profile Measures

Experience with	**Minimum**	**Maximum**	**Mean**
Point and click	5.00	5.00	5.00
Online DBs	3.00	5.00	4.20
WWW searching	4.00	5.00	4.80
Sci Ed research	1.00	5.00	3.20
Video D-Libs	1.00	4.00	2.00

J.2 During-Test Performance Measures

	Task Completion	**Time of task completion**	**Steps**	**Errors**
Overall	74%	4:16	13.40	4.88
Complex Tasks	76%	4:04	12.27	4.93
Combo-Complex Tasks	70%	4:35	15.10	4.80
Task 1	60%	2:02	6.60	3.00
Task 2	70%	3:51	13.40	2.40
Task 3	80%	3:44	8.60	4.60
Task 4	90%	6:27	21.60	7.20
Task 5	70%	5:20	16.80	7.20

J.3 Use of Interface Features and Functions per Task

	Keyword searches	Browse video headings	Browsing of search results	Reexamining search results	Promoted a search result
Overall	4.04	2.08	3.08	0.20	0.24
Complex Tasks	3.93	2.00	2.93	0.33	0.20
Combo-Complex	4.20	2.20	3.33	0.00	0.30
Task 1	0.80	1.80	1.60	0.20	0.40
Task 2	2.60	1.00	4.20	0.00	0.20
Task 3	4.20	1.40	0.60	0.00	0.00
Task 4	5.80	3.80	6.60	0.80	0.20
Task 5	5.80	3.40	2.40	0.00	0.40

J.4 Post-Search Measures

	Familiarity with Search Task, Topic	Satisfaction with Search Results	Adequacy of System Functionality	Representation of Search Task
Overall	3.28	2.92	2.84	3.44
Complex Tasks	3.27	3.20	3.07	3.47
Combo-Complex	3.30	2.50	2.50	3.40
Task 1	3.20	3.60	4.00	3.40
Task 2	3.40	3.00	3.00	3.80
Task 3	3.20	3.80	3.00	3.80
Task 4	3.40	2.20	2.20	3.20
Task 5	3.20	2.00	2.00	3.00

J.5 Post-Experiment Measures

	Minimum	Maximum	Mean
Usability	2.00	4.00	3.00
Learnability	3.00	5.00	3.60

Appendix K

During-Test Results Note Sheet

SUBJECT # **CONDITION #**
TASK # **TASK CATEGORY:**
START TIME: **END TIME:** **TASK TIME:**
NUMBER OF STEPS:

BROWSE	**MORE**	**DETAILS**
KEYWORD	**AND SEARCH**	**BACK**

PROMOTE SEARCHES:

COLOR	**SHAPE**	**TEXTURE**
ALL VISUALS	**HYBRID**	**TEXTUAL**

NUMBER OF ERRORS:

BROWSE	**MORE**	**DETAILS**
KEYWORD	**AND SEARCH**	**BACK**

PROMOTE SEARCHES:

COLOR	**SHAPE**	**TEXTURE**
ALL VISUALS	**HYBRID**	**TEXTUAL**

SUCCESSFULLY COMPLETED? YES NO PERCENTAGE___%

NOTES:

Appendix L

Informed Consent Sheet of Formal Study

Study # 05-10652

INDIANA UNIVERSITY - BLOOMINGTON

A Domain-Centric Approach to Designing User Interfaces of Video Retrieval Systems

You are invited to participate in a research study. The purpose of this study is to evaluate how search tasks and professional domain influence interaction with and effectiveness of video retrieval systems. Results of this experimental study will be analyzed as implications for designing user interfaces of video retrieval systems.

INFORMATION

Participation in this study requires a number of different tasks including:

1. Fill out a pre-test questionnaire covering background and demographics.
2. Search for 6 assigned search topics, or information needs, using the given video retrieval system. Subjects are encouraged to end each search task whenever he/she feels that successful results have been retrieved. Subjects may discontinue any search task at any time. Behaviors and comments about system usefulness will be noted by the experimenter.
3. Fill out a brief post-search questionnaire after the completing each search task.
4. Fill out a post-experiment questionnaire after completing all search tasks. The post- experiment questionnaire will ask questions related to the usefulness and effectiveness of the evaluated system.
5. Answer several questions orally, i.e. a short interview, where responses will be audio recorded.

Audio recordings of post-experiment interviews will be transcribed, coded, and erased within two months of experiment completion. Recordings of the interviews are strictly for research purposes and only the investigators will have access to the content. Audio recordings will not include any reference to the subject's real name or personal identification numbers. Subjects will be allowed to review to their recording at anytime after the conclusion of the experiment. If any subject withdraws from the experiment prior to completion, all audio recordings will be erased immediately. Interviewing questions will focus on the effectiveness of the retrieval system and user interface, not on the subjects themselves.

This experimental session is expected to last approximately 45 to 60 minutes and subjects will be given 5 to 10 minutes per search task. This study will involve 25 to 35 human subjects.

RISKS
There are no foreseeable risks associated with this research.

BENEFITS
This study strives to advance video retrieval research. One benefit is that this study investigates problems surrounding video retrieval from a user-centered perspective, whereas the current focus of video retrieval research is predominately systems-centered. This study also attempts to discover why current video retrieval experiments do not produce significant results by incorporating different methods for systems evaluation.

CONFIDENTIALITY
Subject confidentiality will be protected. At no point in the experiments, questionnaires, or interview will real names or personal identification numbers be recorded. All ensuing reports will discuss subjects using an assigned identifier. Experimental data, including audio tapes, completed questionnaires, interview responses, and experimental notes, will be locked in a secure location where only the investigators will be given access. Identity of the subject can only be broken if the subjects choose to discuss their participation with others. Audio recordings deriving from this study will be destroyed by January 1, 2008.

COMPENSATION
For participating in this study you will receive $10 / hour. If you withdraw from the study prior to its completion, you will receive compensation based on total time spent performing experiments.

CONTACT
If you have questions at any time about the study or the procedures, you may contact the researcher, Dan Edward Albertson II, at Herman B. Wells Library, LI011, 812-360-0579, and daalbert@indiana.edu.

If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact the office for the Indiana University Bloomington Human Subjects

Committee, Carmichael Center L03, 530 E. Kirkwood Ave., Bloomington, IN 47408, 812-855-3067, or by e-mail at iub_hsc@indiana.edu.

PARTICIPATION

Your participation in this study is voluntary; you may refuse to participate without penalty. If you decide to participate, you may withdraw from the study at anytime without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed.

CONSENT

I have read this form and received a copy of it. I have had all my questions answered to my satisfaction. I agree to take part in this study.

Subject's signature________________________________ Date _________________

Appendix M

Search Tasks of Formal Study

1. Find any segment from the NASA Connect Program *Plane Weather.*

2. Find a segment that shows blue skies.

3. Find a student activity that that explores the Fibonacci Sequence, and next a classroom activity that teaches students how to measure shadows and the size of angles.

4. Find two segments that show different types of clouds; one segment that shows fair weather, or cumulus, clouds, and another segment that shows low-level, or stormy, clouds.

5. Find one segment that talks about the surface of Mars; the segment must also show an exploratory robot moving along or setting on the surface.

6. Find one segment that discusses how early pilots navigated while flying, which actually shows older – i.e. black and white – footage of a flying airplane.

7. Find a segment that discusses GPS technologies, and then another segment showing an actual (green) radar screen used by air traffic controllers.

8. Find a segment that discusses airflow and drag of automobiles, then another segment showing a red car involved in a simulated crash test.

Appendix N

Pre-Test Questionnaire of Formal Study

Searcher #___________ Study #05- 10652

INDIANA UNIVERSITY, BLOOMINGTON

What is your highest level of education? Degree______ Major___________ Year ____

Are you currently a student? □ **YES** □ **NO** If **YES**, what is your current major, concentration(s), and year? ____

Do you have an academic background in Science Education or Science Librarianship? □ **YES** □ **NO** If **YES**, which one? ____

Do you have any formal training or professional experience in Science Education or Science Librarianship? □ **YES** □ **NO** If **YES**, which one and how many years experience ____

What is your current occupation? ___ How many years have you been at this job? ____

What levels/grades of students do you primarily teach or assist (or interested in teaching if you're a current student)?

CIRCLE ONE **K** **1 – 5** **6 – 8** **9 – 12** **13+**

What is your gender? □ **FEMALE** □ **MALE** What is your age? ________ years

Please circle the number closest to your experience

How much experience have you had:	**None**		**Some**		**A great deal**
1. Using a point-and-click interface (e.g., Macintosh, Windows)	1	2	3	4	5
2. Searching commercial online systems (e.g. Dialog, Lexis-Nexis, EBSCO, Web of Science)	1	2	3	4	5
3. Searching WWW search engines (e.g., Google, Yahoo)	1	2	3	4	5
4. Searching video retrieval systems or video digital libraries	1	2	3	4	5
5. Searching on any other science educational search systems: specify the system(s) ___________________________	1	2	3	4	5
6. Preparing science education lesson plans, assignments, or any other type of science education projects	1	2	3	4	5
7. Using video as an educational resource (i.e. lab demonstrations, classroom activities, etc.)	1	2	3	4	5

Appendix O

Post-Search Questions of Formal Study

Searcher #____________ Condition ____________ Task # ______________

	None		**Some**		**Extreme**
1. How familiar are you with the subject matter of this search task?	1	2	3	4	5
2. How representative is this task of “real” search tasks encountered by science teachers or librarians?	1	2	3	4	5
3. How easy was it to search on this topic?	1	2	3	4	5
4. Are you satisfied with your search results?	1	2	3	4	5
5. How adequate was system functionality for completing this search task?	1	2	3	4	5
6. How supportive was the user interface for this particular task?	1	2	3	4	5
7. Was it useful to search this topic using visual attributes, i.e. color, shape, texture, all, etc.?	1	2	3	4	5
8. Was it useful to search this topic using keywords?	1	2	3	4	5
9. Was it useful to browse for this topic using Video Titles or Durations?	1	2	3	4	5
10. To what extent do you feel that you completed this search task?	1	2	3	4	5

Appendix P

Post-Experiment Questionnaire of Formal Study

Searcher #____________ Condition ____________

	Not at all		**Somewhat**		**Extremely**
1. How easy was it to *use* this information system?	1	2	3	4	5
2. How easy was it to *learn to use* this information system?	1	2	3	4	5
3. How useful was the keyword search feature in general?	1	2	3	4	5
4. How useful was the video (title) browse feature in general?	1	2	3	4	5
5. How useful was the Promote search in general?	1	2	3	4	5
5A. Usefulness of Color Promote search?	1	2	3	4	5

5B. Usefulness of Shape Promote search?	1	2	3	4	5
5C. Usefulness of Texture Promote search?	1	2	3	4	5
5D. Usefulness of All-Visuals Promote search?	1	2	3	4	5
5E. Usefulness of Hybrid Promote search?	1	2	3	4	5
5F. Usefulness of Textual Promote search?	1	2	3	4	5
5G. Usefulness of attribute, i.e. color and keyword, weighting?	1	2	3	4	5
6. How easy was it to assess video content?	1	2	3	4	5
6A. Is it necessary to view the playing videos?	1	2	3	4	5

Appendix Q

Interview Form of Formal Study

Searcher #____________

What comments do you have about interface features and functions?

How could the interface be more effective?

What specific interface features were most helpful and why?

What did you like/dislike about the search feature? Browse feature?

Are there any interface features that are missing? Any features you would like to see when retrieving video?

What comments do you have about your searching experience or the retrieval system in general?

What did you like about each of the systems?

What did you dislike about each of the systems?

Note questions or comments that the participants raised during search experiment or interview. Record any problems you notice or anything that seems interesting to you about their searching behavior.

DAN EDWARD ALBERTSON II

2757 E. Garden Path
Bloomington, IN 47401
Phone: 812-360-0579
Email: daalbert@indiana.edu
Homepage: http://ella.slis.indiana.edu/~daalbert/homepage/

OBJECTIVE

Employ user-centered strategies for investigating information retrieval and digital library technologies. Develop solutions for information delivery and access for both multimedia and text-based resources. Evaluate domain- and task-specific retrieval systems.

EDUCATION

July, 2007

Ph. D. in Information Science
Ph. D. Minor in Computer Science
Indiana University, Bloomington

Dissertation: A Domain-Centric Approach to Designing User Interfaces of Video Retrieval Systems

Final Defense June 21, 2007

May, 2002

Master of Information Science
Concentrations: Information Retrieval, Human-Computer Interaction
Indiana University, Bloomington

June, 2000

Bachelor of General Studies
Outside areas: Information Technology and Telecommunications
Indiana University, Bloomington

TEACHING EXPERIENCE

Spring, 2004	I502: Information Management Associate Instructor School of Informatics, Indiana University
Summer, 2003	L401: Computer-Based Information Tools - Web Publishing Module Adjunct Lecturer (Two Sections) School of Library and Information Science, Indiana University
Spring, 2003	LIBR202: Information Retrieval Part Time Faculty (Distance Education) School of Library and Information Science, San Jose State University

PROFESSIONAL EXPERIENCE

October, 2004 - Present

Program Manager
Digital Libraries Education Program (DLEP), Indiana University
http://lair.indiana.edu/research/dlib/

This position requires a number of administrative tasks including workshop coordination, project presentations, assisting DLEP fellows and applicants, and website maintenance. I am responsible for reporting to the Institute of Museum and Library Services (IMLS), the funding agency, and serve as the main contact for students and faculty. I am required to be active in a number of academic forums including the Digital Library Federation (DLF) and the Joint Conference on Digital Libraries (JCDL).

January, 2002 - August, 2004

Research Assistant
Laboratory for Applied Informatics Research (LAIR)
Indiana University, Bloomington
http://lair.indiana.edu/research/

Project lead for ViewFinder, a video retrieval research project. Tasks performed for this position include system development and evaluation, and participation in academic forums, such as the Text Retrieval Conference's Video Retrieval Workshop (TRECVID). Systems-centered tasks ranged from backend database development to interface design. Research at LAIR is ongoing for dissertation purposes.

January, 2002 - May, 2002	Research Internship Information in Place, Bloomington, Indiana Assigned to research project that involved a literature survey of augmented reality user interfaces for maritime information systems. Part of a study for U.S. Coast Guard.
May, 2001 - December, 2001	Reference Assistant Undergraduate Library Services (UGLS), Indiana University Assisted library patrons with academic research and using information resources. Provided limited technical support.

PUBLICATIONS

Albertson, D. & Mostafa, J. (2004). TRECVID 2004: Video search experiments at IUB. *Proceedings of the TREC Video Retrieval Evaluation (TRECVID 2004), November, 15 - 16, Gaithersburg, MD.*

Yang, K. & Albertson, D. (2003). WIDIT in TREC 2003 web track. *Proceedings of the 12th Text Retrieval Conference (TREC2003), November, 18 - 21, Gaithersburg, MD.*

Albertson, D., Mostafa, J. & Fieber, J. (2003). Video searching and browsing using ViewFinder: Participation and assessment in TRECVID-2003. *Proceedings of the TREC Video Retrieval Evaluation (TRECVID 2003), November, 17 - 18, Gaithersburg, MD..*

Albertson, D., Mostafa, J. & Fieber, J. (2003). Video Searching and browsing using ViewFinder: Interactive search experiment for TRECVID-2003. *Notebook of the TREC Video Retrieval Evaluation (TRECVID 2003), November, 17 - 18, Gaithersburg, MD..*

Albertson, D., Mostafa, J. & Fieber, J. (2002). Video searching and browsing using ViewFinder. *Proceedings of the 11th Text Retrieval Conference (TREC2002), November, 19 - 22, Gaithersburg, MD.*

Albertson, D., Mostafa, J. & Fieber, J. (2002). Video searching and browsing using ViewFinder. *Notebook of the Text Retrieval Conference (TREC2002), November, 19 - 22, Gaithersburg, MD.*

AWARDS AND HONORS

2005	Award for Service – Received at the 2005 SLIS Doctoral Student Research Forum
2004	School of Library and Information Science Travel Grant
2003	School of Library and Information Science Travel Grant
2002	Ten Month Student Research Fellowship – IUPUI Two Year Student Research Appointment – SLIS-IUB School of Library and Information Science Travel Grant

CONFERENCE / WORKSHOP COORDINATION

2007

ACM-IEEE Joint Conference on Digital Libraries (JCDL)
Digital Libraries Education Workshop
June 18, University of British Columbia, Vancouver

Helped organizing *Developing a Digital Library Education Program*, a workshop hosted by Indiana University, the University of Illinois, the University of North Carolina-Chapel Hill, and Virginia Tech. This workshop was the third and final conference activity supported by IMLS for the current funding cycle.
http://lair.indiana.edu/research/dlib/jcdl07/index.php
http://www.jcdl2007.org/workshops/workshop1.htm

2006

ACM-IEEE Joint Conference on Digital Libraries (JCDL)
Digital Libraries Education Workshop
June 15, University of North Carolina, Chapel Hill

Helped organize *Developing a Digital Library Education Program*, a workshop hosted by Indiana University and the University of Illinois. Responsibilities included: writing publicity for conference website, publicizing workshop, serving as main contact for conference organizers, workshop follow-up, and other administrative duties. This workshop was the second of three conference activities supported by IMLS funds.
http://lair.indiana.edu/research/dlib/jcdl06/index.php
http://jcdl2006.org/program/workshops/#ws-5

2005 SLIS Doctoral Student Annual Research Forum
September 24, 2005, Herman B. Wells Library E174, Bloomington, IN
Chaired conference organization committee.

2005 ACM-IEEE Joint Conference on Digital Libraries (JCDL)
Digital Libraries Education Workshop
June 7, Marriott City Center, Denver, CO

Helped organize *Developing a Digital Library Education Program*, a workshop hosted by Indiana University and the University of Illinois. Responsibilities included: preparing and submitting proposal for workshop, writing publicity for conference website, preparing and submitting paper for proceedings, publicize workshop on various email lists, serve as main contact for conference organizers, and workshop follow-up.
http://lair.indiana.edu/research/dlib/jcdl05/index.php
http://www.jcdl2005.org/workshops.html#0

2005 Digital Library Federation (DLF) Spring Forum
April 13 – 15, Westin Horton Plaza, San Diego, CA

Assisted Kristine Brancolini, Director of the Digital Library Program at Indiana University, with Birds of a Feather Discussion Session covering Digital Library Education.

PUBLIC PRESENTATIONS AND GUEST LECTURES

Spring, 2005 ALA-Student Chapter Meeting
Digital Libraries Education Program at Indiana University
Main Library LI001, Indiana University, Bloomington

Spring, 2004 Digital Library Brownbag Series
TRECVID-2003 and ViewFinder: Research in Video IR
Main Library E174, Indiana University, Bloomington

Spring, 2003 L570: Online Information Retrieval
TREC: Research in IR
Indiana University, Bloomington

Fall, 2002 L570: Online Information Retrieval
TREC: Research in IR
Indiana University, Bloomington

CONFERENCE PARTICIPATION

2006	Association for Library and Information Science Education (ALISE) January 16 – 19, Omni, San Antonio, TX Participated in doctoral student poster session.
2005	Association for Library and Information Science Education (ALISE) January 11 – 14, Hyatt Regency, Boston, MA Promoted Digital Libraries Education Program at Indiana University. Participated in doctoral student poster session. SLIS Doctoral Student Annual Research Forum September 24, 2005, Herman B. Wells Library E174, Bloomington, IN Participated in student poster session.
2004	Text Retrieval Conference’s Video Retrieval Evaluation (TRECVID) November 15 – 16, National Institute of Standards and Technology (NIST), Gaithersburg, MD Participation in TRECVID includes developing video retrieval systems using a common dataset and performing interactive search tasks and evaluation as defined by conference guidelines. Participation also includes presenting findings to other attendees and contributing papers for the conference notebook and proceedings. Institute of Museum and Library Services (IMLS) Outcome-Based Evaluation Workshop December 16 – 17, Washington Terrace Hotel, Washington, DC Participated in workshop to learn responsibilities for measuring and reporting grant progress to IMLS. Required for IMLS funded research projects.
2003	SLIS Doctoral Student Annual Research Forum September 13, Indiana University Memorial Union, Bloomington, IN Participated in student poster session.

Text Retrieval Conference's Video Retrieval Evaluation (TRECVID)
November 17 – 18, National Institute of Standards and Technology (NIST), Gaithersburg, MD

Participation at TRECVID-2003 was similar to the participation at TRECVID-2004 (described above).

2002 Text Retrieval Conference (TREC)
November 19 – 22, National Institute of Standards and Technology (NIST), Gaithersburg, MD

Participation at TREC-2002 was similar to the participation at TRECVID-2004 (described above).

TECHNICAL SKILLS

Programming: Java (7+ years), Perl (5+ years), Assembly

Database: SQL (4+ years), MySQL (1 year), Oracle 8+ and PL/SQL, Microsoft Access

Web and Interface Development: HTML, DHTML, Java Swing, JavaScript

Other Software: Photoshop, Fireworks, Microsoft Office, SPSS

Operating Systems: Unix, Windows, Mac

SERVICE

May, 2005 – July, 2006
Chair
Doctoral Student Association (DSA)
School of Library and Information Science, Indiana University

May, 2005 – July, 2006
Chair
Doctoral Student Research Forum Committee
School of Library and Information Science, Indiana University